(1)

AGARD-AG-308

AGARD-AG-308

AD-A211 106

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE · 92200 NEUILLY SUR SEINE FRANCE

AGARDograph No.308

AMP Working Group 12
and
AGARD Lecture Series 163

Human Performance Assessment Methods

NORTH ATLANTIC TREATY ORGANIZATION

DISTRIBUTION AND AVAILABILITY
ON BACK COVER

89 6 20 175

AGARD-AG-308 (1)

NORTH ATLANTIC TREATY ORGANIZATION

ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT

(ORGANISATION DU TRAITE DE L'ATLANTIQUE NORD)

AGARDograph No.308

AMP WORKING GROUP 12 and AGARD LECTURE SERIES 163

HUMAN PERFORMANCE ASSESSMENT METHODS

The material in this publication was assembled to report the results of AGARD Aerospace Medical Panel Working Group 12 and to support a Lecture Series under the sponsorship of the Aerospace Medical Panel and the Consultant and Exchange Programme. It was presented on 5—6 June 1989 in Downsview (Toronto), Canada, on 12—13 June 1989 in Soesterberg, The Netherlands and on 15—16 June 1989 in Pratica di Mare (Rome), Italy.

THE MISSION OF AGARD

According to its Charter, the mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

- Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community;
- Providing scientific and technical advice and assistance to the Military Committee in the field of aerospace research and development (with particular regard to its military application);
- Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;
- Improving the co-operation among member nations in aerospace research and development;
- Exchange of scientific and technical information;
- Providing assistance to member nations for the purpose of increasing their scientific and technical potential;
- Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

The content of this publication has been reproduced directly from material supplied by AGARD or the authors.

Published May 1989

ISBN 92-835-0510-7

Printed by Specialised Printing Services Limited
40 Chigwell Lane, Loughton, Essex IG10 3TZ

PREFACE

The fighter has always been exposed to stress created by other men and by his physical and mental environment. Stressors usually exert a negative influence on performance, but may also have quite the opposite effect: we have all heard stories of heroic actions, performed under stress, that would be considered impossible under normal circumstances.

Both the commander and the military doctor require knowledge of human response to stressors. The commander must be able to predict the fighting potential of his men; the doctor must be able to offer the appropriate treatment to those whose well-being is jeopardized by stress.

Environmental stressors often produce clear physiological effects, such as cardiac acceleration in the fighter pilot or elevation of core temperature in the tank driver. However, their effects on military task performance remain in many instances rather obscure. Modern sophisticated weapon systems place demands upon the operator's higher mental processes, requiring skills of system management rather than merely of direct psychomotor control. The difficulty of studying these processes is related in part to the intellectual uniqueness of man, which precludes the successful application of animal experimental models. Moreover, psychophysiological methods developed in the laboratory have limited relevance to complex military tasks.

Many research teams have addressed problems of human performance in military environments. However, differences in protocol, data collection, or the conditions of testing have often precluded direct comparison of results, and time, energy, and money have been wasted. The reason for this 'Tower of Babel' phenomenon has been a communication problem created by differences not in national but in scientific language.

Confronted with this exasperating situation, a number of researchers in the NATO member countries met with the intention of introducing a more systematic approach to performance testing. This "Aachen Academic Group", which comprised workers from universities, military establishments, and industry, held a series of meetings sponsored initially by the USAF European Office of Aerospace Research and Development (EOARD) and later by the European Community (EC). Professor Andries F. Sanders received funding from the USAF to conduct a survey of current performance researchers, and reported widespread enthusiasm for the notion of standardization. Subsequently, the AGARD Aerospace Medical Panel formed Working Group 12, whose major tasks were to construct a standardized test battery and to define a data exchange format.

As a first step in promoting the collaboration necessary for a successful programme of standardization, the members of Working Group 12 compiled and published an international register of performance research. The register, although not exhaustive even within the NATO member countries, revealed extensive use of performance tests for a variety of purposes, and it quickly became apparent that the scope of the working group's activities should be delimited. The major applications of performance tests were found to be within the fields of personnel selection and of stress research. Since the former was an extremely wide-ranging topic that had already been considered by RSG 14 of NATO Panel VIII, the efforts of the working group were directed primarily towards the latter. Although the group will make no specific recommendations concerning the application of the standardized battery to selection, it is hoped that the development of a normative data base will be of interest to selection researchers.

The working group set out not to develop new performance tests but to formalize the protocol of tests with a proven record of success in stress research. To ensure maximum generalizability, laboratory-based tests were chosen in preference to simulations of specific real-life tasks. The importance of occupational validity was not ignored, however, and a test was considered suitable for consideration only if there was preliminary evidence for its relevance to performance on practical tasks

Although our objectives may seem limited, the encouragement of closer cooperation between laboratories, the enhancement of comparability between studies, and the definition of a data exchange format will have many potential benefits. Duplication of effort, previously wasteful, will now be used to increase the power of performance tests; the effects of a wide variety of environmental conditions on a particular mental process will become apparent; previously undiscovered patterns of relationships between variables may be revealed; and it may be possible eventually to establis[illegible]nal centralized data base. Thus, the test battery described here, although not immediately providing fresh insights into th[illegible] of human performance, will serve as a framework for the systematic accumulation of knowledge.

Medecin en Chef G.SANTUCCI
Chairman
AGARD AMP Working Group 12

PREFACE

Le combattant a toujours été un homme soumis à des agressions de la part d'autres hommes et à celles dues à l'environnement physique et psychique. Ces éléments stressants ont en général un impact négatif sur les performances humaines, mais ils peuvent avoir également un effet tout à fait inverse et tout un chacun connait des anecdotes, ou des faits dits héroïques, où l'on a vu des hommes effectuer des tâches considérées comme impossibles dans des circonstances ordinaires.

Les chefs et les médecins militaires ont besoin d'informations sur la réponse humaine aux éléments stressants. Le chef doit connaître son potentiel de combat et le médecin doit pouvoir offrir des thérapeutiques adéquates à ceux dont le bien-être se trouve menacé par le stress.

Les agressions ayant pour origine l'environnement physique ont souvent des effets physiologiques très marqués, tels l'accélération du rythme cardiaque subie par les pilotes des avions de combat ou les modifications de température interne des tankistes. Ceci nonobstant, leurs effets sur l'exécution de tâches d'ordre militaire sont loin d'être clairs dans de nombreux cas. Les systèmes d'armes modernes évolués exigent une grande activité intellectuelle de la part de l'opérateur, et ceci dans le domaine de la gestion de systèmes plutôt que dans celui du simple contrôle psychomoteur. Le problème qui se pose pour l'étude de ces processus s'explique en partie par la nature unique des capacités intellectuelles de l'homme, ce qui rend impossible l'emploi de modèles expérimentaux animaliers. En outre, les méthodes psychophysiologiques développées en laboratoire ne sont que partiellement applicable au tâches militaires complexes.

Bon nombre d'équipes de chercheurs ont déjà abordé les problèmes soulevés par l'étude des performances humaines en situation opérationnelle mais très souvent il s'est avéré impossible de faire la comparaison directe des résultats en raison des différences qui existent dans les protocoles, la collecte des données et les conditions d'essais. Beaucoup de temps, d'énergie et d'argent ont été ainsi perdus. Ce phénomène de "Tour de Babel" a pour origine un problème de comunication créé par des différences non pas dans la langue nationale mais dans la langage technique.

Face à cette irritante situation un certain nombre de chercheurs membres des pays de l'OTAN se sont réunis pour tenter d'élaborer une approche plus méthodique des tests de performance. Ce groupe, le "Groupe Académique d'Aachen" composé de chercheurs de tous horizons (universités, institutions militaires, industries, etc.) s'est réuni à plusieurs reprises, d'abord sous l'égide de l'"European Office of Aerospace Research and Development" (EOARD) et ensuite sous celle de la Communauté Economique Européenne (CEE). Une étude fut conduite sous contrat de l'US Air Force par le Professeur Andries F.Sanders auprès des chercheurs travaillant dans ce domaine, afin d'évaluer l'utilité d'un travail qui consisterait à standardiser une batterie de tests de performances mentales humaines. L'accueil de cette démarche par la communauté scientifique fut enthousiaste. C'est ainsi que la Commission de Médecine Aérospatiale de l'AGARD décida la création du Groupe de Travail No.12. La mission du Groupe fut d'élaborer une batterie standardisée de tests et de rechercher et créer une structure pour l'échange de données.

La démarche initiale adoptée par le Groupe de Travail No.12 en vue d'encourager la coopération nécessaire à la réussite d'un tel programme de standardisation, fut de compiler et de publier un Annuaire international des équipes de recherche en performances humaines. Cette publication, quoique loin d'être exhaustive en ce qui concerne les pays membres de l'OTAN, souligne l'emploi très généralisé de tests de performance dans de nombreux domaines. Le Groupe de Travail en a conclu très rapidement qu'il devait délimiter l'étendue des travaux envisagés. Il s'est avéré que les applications principales des tests de performance se trouvaient dans les domaines de la sélection du personnel et de la recherche portant sur le stress. Le premier étant un vaste sujet, déjà examiné par une Groupe d'Etudes et de Recherches de la Commission VIII du Groupe de Recherche pour la Défense de l'OTAN, les efforts de notre Group de Travail ont porté principalement sur la recherche sur le stress. Bien que le Groupe n'ait pas à se prononcer sur les applications de la batterie standardisée de tests pour la sélection du personnel, il est à espérer que le développement d'une base normative de données éveillera l'intérêt des chercheurs dans le domaine de la sélection.

Le Groupe de Travail s'est donné pour but non pas de développer de nouveaux tests de performance mais de formaliser le protocole des tests dont l'efficacité a été confirmée par les spécialistes en la matière.

Des tests de laboratoire ont été choisis de préférence à la simulation de tâches spécifiques réelles, afin de faciliter l'adoption généralisée de ces procédures. L'importance de la validation professionnelle n'a pourtant pas été oubliée. Le critère retenu pour la prise en considération d'un test est la justification préalable de sa pertinence au regard de la performance humaine impliquée dans le travail étudié.

Bien que les objectifs puissent paraître limités, l'encouragement en vue d'une collaboration plus étroite entre laboratoires, l'obtention d'une meilleure comparabilité entre études et la definition d'une structure d'échange de données ne pourront avoir que de conséquences bénéfiques. La duplication des efforts, source de gaspillages dans le passé, servira désormais à renforcer l'efficacité des tests de performance. Les effets de tout un éventail de conditions ambiantes sur un processus mental donné deviendront évidents. Des modèles de relation entre variables, inconnus jusqu'ici, risquent d'être découverts et la création d'une banque de données officielle centralisée pourra s'avérer possible à terme.

Ainsi, la batterie de tests décrite ici, si elle n'offre pas, pour l'instant, de nouvelles élucidations sur la nature de la performance humaine, servira de cadre pour le recueil systématique des connaissances.

Médecin en Chef G.SANTUCCI

AEROSPACE MEDICAL PANEL OFFICERS

Chairman: Mr C.Bates, Jr
Director
Human Engineering Division
AAMRL/HE
Wright-Patterson AFB
Ohio 45433-6573
USA

Deputy Chairman: Dr G.Santucci
Chef du Laboratoire Central de
Biologie Aérospatiale
E.A.S.S.A.A.
5, Bis Av. de la Porte de Sèvres
75996 Paris Armées
France

PANEL EXECUTIVE

Major J.A.Winship
AGARD-NATO
7, Rue Ancelle
92200 Neuilly sur Seine
France

LIST OF CONTRIBUTORS

Working Group Members

*Dr Guy Santucci (Chairman)
C.E.R.M.A.
26, Boulevard Victor
75996 Paris Armées
France

*Dr Eric Farmer
RAF Institute of Aviation Medicine
Farnborough, Hants GU14 6SZ
United Kingdom

*Dr James D.Grissett
Naval Aerospace Medical Research Lab.
Pensacola, Florida 32508
USA

*Dr A.Wetherall
Chemical Defence Establishment
Porton Down
Salisbury, Wilts SP4 0JQ
United Kingdom

*Dr Louis Boer
Institute for Perception TNO
Kampweg 5
3769 ZG Soesterberg
The Netherlands

*Dr K.M.Goeters
DLR Dept. of Aviation and
Space Psychology
Sportallee 54
D-2000 Hamburg 63
Germany

*Mr Edwin Schwartz
DLR
Institute für Flugmedizin
Postfach 90 60 58
5000 Köln 90
Germany

*Dr Glenn Wilson
AAMRL/HEG
Wright-Patterson AFB
Ohio 45433-6573
USA

Secretary to the Working Group

Mr R.E.Yates
AAMRL/HEG
Wright-Patterson AFB
Ohio 45433-6573
USA

* Faculty Member — AGARD AMP Lecture Series 163.

Consultants

Psychometrics

Prof P.Kline
Department of Psychology
Washington Singer Laboratories
University of Exeter
Exeter, EX4 4QG
United Kingdom

UTC-PAB

Lt D.L.Reeves
Naval Aerospace Medical Research
Laboratory
Pensacola, Florida 32508
USA

Data Base

Mr G.W.Hart
Technical Information Panel
AGARD
7, Rue Ancelle
92200 Neuilly sur Seine
France

CONTENTS

CHAPTER 1

INTRODUCTION

A. PERFORMANCE TESTING AND THE NEED FOR STANDARDIZATION

There is growing interest in the effects of environmental stressors on human performance. Particular attention has been given to military and industrial tasks in which stress-induced error may have serious consequences. Unfortunately, differences in testing procedures have hindered the integration of findings for a particular task or a particular stressor.

In traditional psychometrics, a lengthy development phase usually precedes presentation of the test in a completely standardized form. However, the performance tests used in stress research are often borrowed from techniques reported in the theoretical literature on human cognition. These techniques take the form of paradigms within which specific variables are manipulated experimentally. Consequently, no standard protocol is available, and it is unsurprising that applied researchers construct versions of the test that, although conforming to the paradigm, differ considerably in detail.

Sternberg's (1966) memory search technique is an example of a performance test that was originally developed as a theoretical tool. A 'memory set' of items is presented, followed by a 'probe' item, and the subject is required merely to indicate whether the probe was present in or absent from the memory set. Despite the simplicity of this procedure, considerable variation is possible. For example, the memory set may be fixed or variable; the range of memory set sizes may vary; the inter-stimulus interval may be experimenter-or subject-paced; and the stimuli may be familiar or unfamiliar.

Systematic variations within the memory search paradigm, such as the use of visually degraded probes, are of great interest to the theorist. Applied researchers, however, require of any test that it serves as a constant yardstick against which to measure the effects of variation in the environment. Within a single experiment, this objective is easy to attain: the same version of the test can be administered under different environmental conditions, and, provided that a sound experimental design has been employed, any differences in performance can be attributed to the environment. However, problems emerge when an attempt is made to integrate findings from different laboratories. Variations in test protocol represent a source of confounding, and preclude direct comparison of results.

Well-accepted paradigms such as memory search form the building blocks for test batteries that provide broad profiles of human performance. Such batteries are usually developed in response to an applied problem such as selection for employment or evaluation of the effects of an environmental stressor on job performance, and represent an attempt to solve the applied problems of a particular sponsor. Sanders, Haygood, Schroiff, and Wauschkuhn's (1986) survey of performance test batteries, and the discussions of performance researchers comprising the 'Aachen Academic Group', indicated a surprising degree of consensus in the selection of tests. The Aachen Group concluded that a core of commonly used performance tests could be selected for inclusion in a standardized battery, and that a normative data base, comparable to that available for intelligence and personality tests, could then be established.

Working Group 12 of the AGARD Aerospace Medical Panel was formed to achieve this objective. To facilitate communication between researchers, the working group initially compiled an international register of performance research (AGARD Report No. 763), which included details both of tests and of applications. Seven common paradigms, each with preliminary evidence of psychometric soundness, were selected as the basis of the AGARD Standardized Tests for Research with Environmental Stressors (STRES) Battery.

The AGARD STRES Battery can be considered an extension of the approach initiated by representatives of the US Navy, Air Force, and Army in the development of the Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB). The UTC-PAB is designed to be a dynamic system that will evolve through several stages; it provides the option to use a core subset of tests or to construct a unique combination of UTC-PAB tests to meet specific requirements (see Englund, Reeves, Shingledecker, Thorne, Wilson, & Hegge, 1987). The STRES Battery places even greater emphasis upon standardization. It represents the collaborative efforts of an international group of users to define the tests most useful in a battery for stress research, provide detailed and machine-independent test specifications, and establish a standardized data exchange format to facilitate the construction of a data base. To ensure maximum applicability, language differences have been taken into account.

The benefits of this standardization programme include the opportunity to apply both 'narrow-band' and 'broad-band' strategies (Hockey & Hamilton, 1983) to stress research. The narrow-band approach involves examination of the effects of a variety of stressors on performance of a single task, and permits generalizations concerning the effects of stressors; the broad-band approach, in which the effects of a single stressor on various tasks is investigated, helps to reveal subtle but important differences between stressors. Data exchange will also permit examination of the effects of incidental variables such as age and sex on test performance, and the inclusion of occupational information may permit application to personnel selection.

The AGARD STRES Battery is intended to inhibit neither the systematic manipulation of test variables that is of central importance in theoretical research, nor the generation of new approaches to performance testing. Rather, its objective is to provide a solid core of well-accepted performance tests for use by the applied researcher.

B. APPLICATIONS OF HUMAN PERFORMANCE TESTING

There are two broad classes of purpose for a battery of performance tests. It can be used to evaluate the effects of environmental stressors, or to assess the information-processing abilities of individuals. To evaluate stressor effects, emphasis is placed upon comparison of the performance of groups of subjects under control conditions to that under unfavourable conditions such as sleep loss and fatigue; monotony and boredom; illnesses; toxic fumes; hypoxia; and alcohol and other drugs. The ultimate goal is to assess the extent to which a particular stressor influences performance in real-life situations. In the assessment of abilities, on the other hand, interest lies in differences between individuals. This application is comparable to classical test psychology. The individual's score is used as a measure of information-processing capability relative to that of other individuals. Both applications depend upon the assumption that it is possible to generalize from performance on laboratory tasks to that on practical tasks; in other words, that the variance of the performance measure is not test-specific but relates to real life.

The AGARD STRES Battery is concerned primarily with stress research, the requirements of which differ in some respects from those of ability assessment. To assess individuals, test measures should ideally be relatively insensitive to variations in environmental conditions but sensitive to individual differences. To assess stressor effects, the opposite is true: performance should fluctuate markedly when environmental conditions change, but the variance due to individual differences should ideally be small. Figure 1 provides an illustration of both types of task.

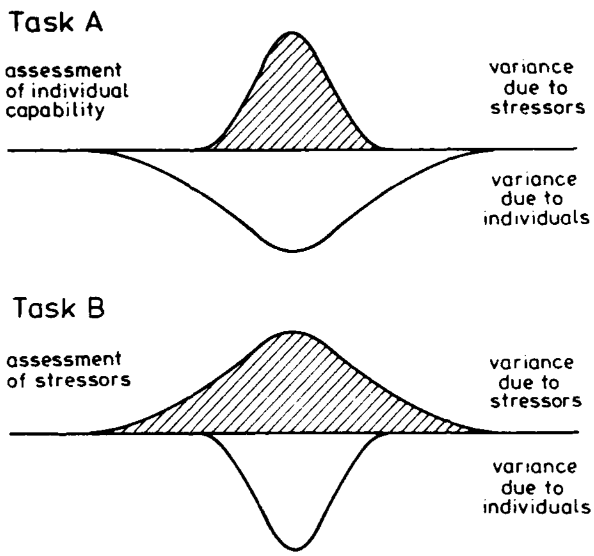


Figure 1. Differential sensitivity to stressors and individual differences. Task A is more sensitive to individual differences; Task B is more sensitive to stressors.

In practice, a test may be found to be sensitive both to stressor effects and to individual differences, and for this reason the potential application of the STRES battery to personnel selection will not be ignored.

C. HUMAN PERFORMANCE THEORY: SCOPE AND LIMITATIONS

The STRES Battery is not dependent upon a specific theoretical standpoint. Nevertheless, it is necessary to consider the general nature of models of human performance, the mental processes that commonly used performance tests purport to measure, and the ways in which these tests differ from real-life activities.

The aim of Human Performance Theory (HPT) is to search for lawful relations between task variables and performance. This has led to the development of a large number of information-processing models. Despite the differences between competing approaches, it is relatively straightforward to extract common assumptions and ideas, and hence to arrive at a 'modal model' of the organization of the human information-processing system.

The central assumption underlying most models is that man is a single information-processing system equipped with memory stores, or an ensemble of such systems each with its own functional significance. This so-called computer analogy incorporates the notion of limited capacity, which suggests both that mental processes are time-consuming and that the time required increases with complexity. Thus 'mental chronometry', in which mental processes are investigated by dissection of reaction time (RT), is one of the most important tools of the performance theorist.

A very general information-processing model is that of the Perception-Decision-Action (PDA) cycle shown in Figure 2. Perception and action are the input and output functions, respectively, with decision as the intervening process. Figure 3, which shows the various stages of the reaction-time process in addition to some of the task variables affecting these stages, can be considered a more specific elaboration of the PDA cycle.

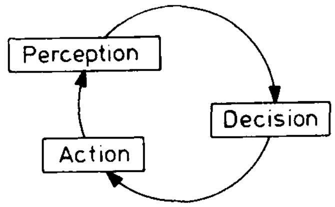


Figure 2. The Perception-Decision-Action model.

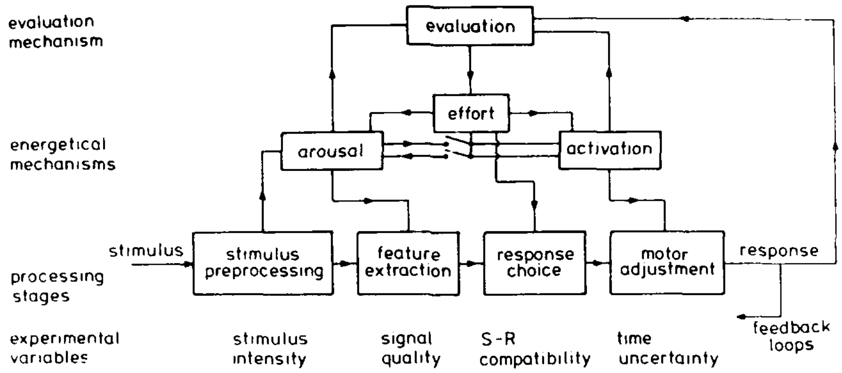


Figure 3. Energy and structure. The model shows the structure of the reaction process (bottom line), and energetical supply to these structural elements. (Adapted from Sanders, 1983.)

In this model, the structural properties of information processing have been expanded by taking into account the dimension of energetical supply. The supply to perceptual structures is called arousal, and that to motor-related structures is called activation. The concept of energetical supply, or amount of mental resources available to the information processing structures, is very important in the present context of stress research.

Sufficient resources for adequate task performance are normally allocated to processing structures with little conscious effort. Stressors or suboptimal conditions, however, may hinder the supply of resources, either by reducing the total amount of energy, or by directing the flow of energy to activities unrelated to, or even detrimental to, adequate task performance. Energy reduction has been postulated to occur under conditions of fatigue, boredom, and sleepiness; energy diversion under conditions related to anxiety and worry. As a consequence, tasks are not always provided with the necessary resources, and information-processing performance will suffer. The extent of performance deterioration can be taken as an indication of the effect of the stressor.

Stressors such as diazepam may have a relatively short-lived effect on performance (see Figure 4a). Conversely, other stressors may fail to produce performance degradation during the first few minutes of testing. The initial challenge and stimulation provided by the performance test may, for example, be sufficient to counteract the effects of sleep loss for the first 5-10 minutes (see Figure 4b). Indeed, in some studies an uninterrupted testing period of 20-30 minutes is necessary to demonstrate degradation. The duration of tests in the STRES Battery may be increased where appropriate, using multiples of the recommended value.

It is important to recognize that inferences about the effects of stressors have an indirect quality. For example, one of the effects of fatigue is a deterioration of information processing (see Figure 4b for an illustration). It is therefore quite legitimate to suggest that performance tasks can be used to measure fatigue. However, one should bear in mind that mental performance may also be affected by other stressors, by differences in individual capability, and by amount of practice. A thorough knowledge of the situation is therefore essential to demonstrate unequivocally that the deterioration in performance is attributable to fatigue. For this reason, investigators try to manipulate the stressor of interest but to eliminate confounding due to other stressors, individual differences, and practice. Interpretation of mental performance is possible only in such controlled environments.

As discussed earlier, the STRES battery is a sample from the paradigms developed in HPT, many of which depend upon measurement of the time between presentation of a target stimulus and execution of a pre-defined response. In theoretical research, task parameters are typically varied, in an otherwise constant environment, to extract general principles of human performance; in applied research, however, task parameters are generally held constant in a changing environment, to discover the effects of external factors on performance. HPT paradigms have been used, for example, to establish that performance

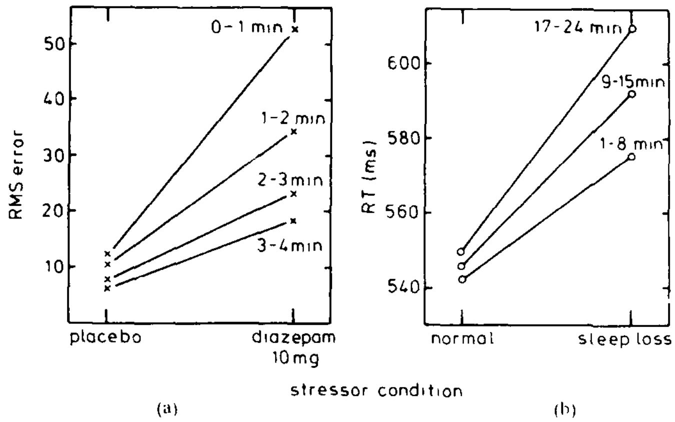


Figure 4. Performance as a function of stressor condition, with time on task as the parameter. In the left panel, the effect of diazepam on tracking performance declines during the first four minutes of testing; in the right panel, the effect of sleep loss on RT increases as a function of time on task.

declines with age; that brain damage, minor illnesses such as colds or influenza, and emotional disturbances such as depression and anxiety, all have adverse effects on information processing; and that personality characteristics such as extraversion influence performance.

Clearly, the tightly-constrained paradigms of HPT sample only a subset of human behaviour. In the following discussion, an attempt will be made to place laboratory performance in a proper perspective, by considering the dimensions of complexity and hierarchy in human information processing. These dimensions may be correlated, since in many situations complex processing will be associated with high levels in the hierarchy, whereas low hierarchical levels will be associated with more simple and well-defined tasks; such a relationship, however, is not inevitable.

Complexity of information processing is determined by the nature of the stimuli, the rule by which stimuli are mapped to responses, and the type of response required. HPT paradigms use only highly-structured information-processing tasks. Stimuli are well-defined units such as letters, words, or tones; responses are key-press reactions or vocal utterances; and stimulus-to-response (S-R) mappings are unambiguously specified. Moreover, the tasks have well-defined starting and end points.

Note that the clear definition of the stimulus does not imply easy identification. In vigilance tasks, for example, a signal defined as a tone of exactly 0.8 seconds' duration may be embedded within a sequence of non-signals of 0.7 seconds' duration. However, all HPT tasks exclude the ambiguity sometimes encountered in real-life activities, in which the individual must determine the exact nature of the situation before deciding whether action should be taken and, if so, what type of action is required. They also exclude unusual and unexpected events to which novel responses must be generated.

Real-life stimuli may be more complex than those used in HPT paradigms. They may comprise many different elements, perhaps requiring temporal integration over long periods of time; they may be hidden or masked by other meaningful stimulus patterns; and they may occur unexpectedly. At extremely high levels of complexity, the classification of stimuli may represent a source of contention even among experts. Examples include medical diagnosis based on subjective complaints, medical examination, and laboratory analysis, and the problem of identifying and interpreting political or economical emergencies.

Real-life responses and S-R mapping rules may also be more complex than those of HPT paradigms. The selection of an appropriate response may require consideration of factors ranging from conventional wisdom to economic necessities and social or political consequences. It may be necessary to discriminate between many possible courses of action, and this process may take much longer than is permitted by any laboratory task. Alternatively, the problem may require the creation of an entirely novel type of response, a 'divergent' solution rather than the 'convergent' solution required by performance tests.

The typical HPT task presents a repetitive succession of very similar but discrete S-R cycles. A real-life task, on the other hand, may comprise a single S-R cycle. Moreover, real-life tasks may lack well-defined starting or end points, and may have cumulative aspects in which task difficulty depends on past performance. In most performance tests, with the possible exception of continuous tracking, fatigue and practice produce the only cumulative effects.

The second dimension for a proper perspective on HPT is hierarchy of processing, in which Perception-Decision-Action cycles occur at different levels. In a hierarchical task, higher levels initiate lower levels, and lower levels influence higher levels by providing them with feedback concerning their outcomes. Lower-level processes can be changed or interrupted by higher levels; such changes or interruptions can be understood only from the perspective of the higher level, not from observing the lower levels in isolation. The hierarchical nature of behaviour was emphasised by Miller, Galanter, and Pribram (1960), who argued that even simple activities such as hammering a nail into a piece of wood could be characterised as a hierarchy of TOTE (Test-Operate-Test-Exit) units. An analogy can be drawn with military operations, in which high level strategy determines the choice of tactics, and may then be modified by the outcome of these tactics.

In summary, it is apparent that the focus of HPT is on the mechanisms of information processing rather than the influence of environmental, social, emotional, or personality factors. Nevertheless, standardization of a particular HPT technique can produce a test suitable for assessing the effects of environmental change. Since such a test depends upon tightly constrained domains of stimuli and responses, and samples relatively low-level behavioural cycles, it is most relevant to well-defined real life tasks. The activities of the aircraft pilot, for example, can be divided into sub-tasks that resemble HPT-derived tests. When controlling attitude, the pilot must extract signals concerning the position of the horizon, and make relatively simple manual corrections. On the other hand, some practical tasks bear little obvious relationship to the mental processes measured by traditional performance batteries. For example, the complex decision processes required of the military commander are not well represented by performance tests requiring specific responses to well-defined stimuli. In general, these tests are more easily applicable to human performance in man-machine systems than to decision-making.

CHAPTER 2

THE AGARD STRES BATTERY

A. FUNDAMENTALS OF PSYCHOMETRICS AND EXPERIMENTAL DESIGN

Psychological tests must satisfy certain psychometric criteria. Moreover, they must be used within a sound experimental design. The following notes are included for the guidance of those who wish to apply the AGARD STRES Battery to stress research, but who have limited experience of psychological testing.

Psychometric principles

Any psychological test must exhibit the properties of validity, reliability, and sensitivity. In other words, it must measure what it purports to measure, do so consistently, and be capable of detecting the effects of the environment or of individual differences in ability.

It is sometimes suggested that high reliability is undesirable in a test designed to exhibit intra-individual variability under the influence of environmental conditions. However, this is to confuse the notions of reliability and sensitivity: the former is concerned with the amount of error variance in test scores, whereas the latter refers to variation induced by environmental change. Thus, the test should have high test-retest reliability, indicating stability under constant testing conditions, together with high internal reliability, but it should reflect changes in variables to which it is designed to be sensitive.

The reliability of a performance task may be affected by practice. Generally, task performance improves systematically until an asymptotic level is reached, although additional but more subtle improvement may occur in the form of overlearning (or, in more modern terminology, a transition from 'controlled' to 'automatic' processing), during which the amount of mental resources required to perform the task declines. The specification of each STRES task includes both a standard and an abridged training schedule. It is strongly recommended that the standard schedule be adopted, to ensure that most of the effects of practice are eliminated prior to the experimental phase. The abridged schedule may be used if practical constraints limit the time available for testing. Since, however, some effect of practice is likely to be observed during the experimental phase, particular attention must be paid to balancing the order of conditions.

The available evidence suggests that, after training, STRES task scores will achieve an acceptable level of reliability. High reliability is a necessary, but not a sufficient, condition for high validity. In other words, the target attribute cannot be measured adequately by a test that fails to provide consistent scores, and may not be measured adequately even by a reliable test. Validation is therefore an essential component of the development of the STRES battery.

Construct validity is important in the present context, since it indicates the extent to which performance is consistent with theoretical predictions concerning the nature of the mental process that the tests are designed to measure. Approaches that will be adopted to investigate this and other aspects of the validity of the STRES battery are outlined in Chapter 4.

The existing evidence of reliability, validity and sensitivity is reviewed in the specification of each STRES test. In most instances, this information is incomplete. Only the adoption of standardized test protocols will permit rigorous investigation of the psychometric properties of the tests.

Experimental design

Any assessment of performance must obviously be conducted under carefully controlled conditions. An 'independent' variable is manipulated systematically to discover its effect on a 'dependent' variable. In the present context, the major independent variables, or 'factors', are stressors or stressor levels, and the dependent variables are response measures provided by the STRES Battery (see Figure 4).

Confounding

It is essential that variation does not occur simultaneously on two or more factors. For example, if, in a study of the effects of noise, males were tested in quiet conditions and females were tested in noise, no conclusions could be drawn concerning the source of performance differences between conditions, since confounding would exist between the factors of sex and noise.

There are several solutions to this problem. For example, sex can be considered a nuisance variable and simply balanced in each condition; or sex can be included as a factor and combined factorially with noise level (each sex performing in both quiet and noise).

Interactions

If more than one experimental factor is present, the data should be analysed using a statistical technique such as analysis of variance (ANOVA), which partitions the total variance in test scores into its separate sources. ANOVA permits the investigation both of main effects (eg the overall difference between the performance of males and females regardless of noise levels) and of interaction effects (variation in the effect of one factor, such as noise, at different levels of another factor, such as sex).

The possible presence of interaction effects must be taken into account during the construction of the experimental design. Consider a hypothetical experiment in which sex has simply been balanced between conditions of noise and quiet. If noise improved the performance of one sex but degraded the performance of the other, the experimenter might erroneously conclude that noise had no effect on performance. Clearly, important variables that may interact with the stressor under study should be included as factors.

Within-and between-subjects designs

Consider a study designed to investigate the effects of a single night's loss of sleep. In its simplest form, the experimental design would comprise two conditions: a control condition in which subjects are tested after a normal amount of sleep, and an experimental condition in which subjects are tested after loss of one night's sleep.

One of the major design issues concerns whether each subject should perform in both conditions (within-subjects design), or whether separate groups of subjects should be tested in each condition (between-subjects design). The within-subjects solution is often favoured because each subject acts as his own control, reducing the possibility of confounding due to pre-existing differences between subjects, and for the practical reason that fewer subjects need be enlisted.

If a within-subjects design were used in which all subjects were tested first in the control condition and then after sleep loss, the beneficial effects of practice might mask the detrimental effect of loss of sleep. The conventional solution to this problem is to balance the order of conditions between subjects (the 'AB-BA' design); in the present example half of the subjects would be assigned to the control condition first and half to the sleep loss condition first. However, this design is based upon the assumption that the transfer between conditions is symmetrical, ie that the effect of practice between the first and second conditions is identical regardless of the order in which the conditions are administered. Unfortunately, there is sometimes clear evidence for asymmetical transfer effects (Poulton & Freeman, 1966). It may be found, for example, that initial performance under stress leads to the adoption of inappropriate methods of completing the performance test that are carried over to the subsequent control condition, whereas initial performance under control conditions produces an efficient strategy that permits performance to be maintained even under stress. When a within-subjects design is used, therefore, the effect of condition order should be examined for possible asymmetrical transfer.

Effects of expectation

Human performance may be influenced by the individual's expectations concerning the effects of stressors. Although ethical considerations demand that subjects be pre-informed of the nature of the stressors to which they are to be exposed, the experimenter should not, if possible, reveal to subjects the order in which the control and experimental conditions are administered. Truly single-blind conditions can be achieved in some drug studies by means of a placebo, but not in studies of stressors such as heat or noise that can be sensed directly by the subject.

B. CRITERIA USED IN THE SELECTION OF THE TESTS

The survey conducted by Sanders et al (1986) was used initially to identify general classes of test that were in common use and that together would provide measures of a wide range of mental processes. Individual tests were then selected on the basis of the following criteria:

1. Preliminary evidence of reliability, validity, and sensitivity.

2. Documented history of application to assessment of a range of stressor effects.

3. Short duration (maximum of three minutes per trial block).

4. Language-independence.

5. Sound basis in HPT

6. Ability to be implemented on simple and easily-available computer systems.

C. TESTS SELECTED FOR THE BATTERY

The following seven tests were selected on the basis of these criteria:

Reaction time

Several reaction time tasks satisfy the criteria listed above. The task selected was based on that appearing in the TNO Taskomat Battery (Boer, Gaillard, & Jorna, 1987), since it provides separate measures of the stages comprising the reaction process.

Mathematical processing

Numerical ability has repeatedly been identified as a factor in factor-analytic studies of skilled performance. Several mathematical processing tasks exist, but most require a numerical response. The Mathematical Processing task from the USAF Criterion Task Set (CTS) and the UTC-PAB was chosen since its two-choice response is more suitable for computerized presentation. This task measures the ability to manipulate arithmetical information, and so places demands upon working memory.

Memory search

Many paradigms exist to investigate aspects of human memory. The Sternberg memory search paradigm was selected because of its popularity in applied performance studies and its ability to indicate the loci of stressor effects.

Spatial processing

Spatial processing tests exist in a variety of forms, some requiring complex hardware. The CTS/UTC-PAB version, which taps visuospatial short term memory by requiring the subject to imagine rotations, was selected because of the well-documented history of application of this general technique, and its ability to be administered using relatively simple hardware.

Unstable tracking

Tracking places demands primarily upon motor-related resources. Of the many tracking tests available, the CTS/UTC-PAB version was selected because of its previous application to stress research, and its sound theoretical basis.

Grammatical reasoning

Some researchers have argued that mathematical and verbal reasoning tasks sample the same resource. However, it has been reported that performance on these two types of test can be differentially affected by some stressors, including drugs (eg Holland, Kemp, & Wetherell, 1978). Both types were therefore included in the present battery.

The STRES grammatical reasoning task requires the manipulation and comparison of grammatical information. It was based on that described by Baddeley (1968), which has been used extensively to measure stressor effects. However, it was necessary to modify Baddeley's method to ensure language independence. Specifically, the use of the passive voice was avoided, since this construction is rarely used in German. To compensate for the consequent reduction in difficulty, the number of statements within each problem was increased.

Dual-task performance

Division of attention between task components is an important element of many practical tasks such as flying, and there is evidence that the allocation of mental resources is affected by stress. It was therefore considered essential to include in the battery a measure of dual-task performance.

Since dual-task performance can be interpreted only in the light of performance on each task in isolation, the total administration time of the battery was reduced by combining two of the tasks already included in the battery. Tracking and memory search were selected because of their relevance to continuous control tasks, such as flying, in which there are periodic demands upon working memory.

D. GENERAL SOFTWARE PARAMETERS

Each STRES task is designed for computerized administration. It is recommended that an overall controlling programme be created to perform the following operations:

i) Request subject information: the information that is required for the data base (see Chapter 3, Section C) should be entered.

ii) Present the tasks in the following, fixed, order:

1. REACTION TIME
2. MATHEMATICAL PROCESSING
3. MEMORY SEARCH
4. SPATIAL PROCESSING
5. UNSTABLE TRACKING
6. GRAMMATICAL REASONING
7. DUAL-TASK (UNSTABLE TRACKING WITH CONCURRENT MEMORY SEARCH)

The programme controlling an individual task should perform the following functions:

i) Present standardized instructions on screen of computer monitor.

ii) Present stimulus sequence according to test description.

iii) Store condition information and performance data on computer disk.

Each task specification includes a detailed description of its parameters and administrative protocol. A flow diagram is included to facilitate the translation of the specification into computer code.

E. *GENERAL CONDITIONS OF TESTING*

The recommendations presented below should be followed as closely as possible. Deviations, where necessary, should be recorded with the experimental data.

Stimulus display

Display elements should be presented in white on a dark background; the ratio of display element to background luminance should be between 7:1 and 12:1. Alphanumeric characters should subtend a vertical visual angle of 15-20 minutes of arc, which, at the recommended viewing distance of 0.6 metre, corresponds to a character height of 2.6-3.5 millimetres. Because of the test battery's dependence upon presentation of visual material, it must be ensured that subjects have normal or corrected-to-normal vision.

Response devices

To run the tests comprising the STRES battery, four response keys and a joystick are required.

Depression of a response key should, where appropriate, cause RT to be recorded to the nearest millisecond. Non-latching, push-to-make switches should be used, with a travel of three millimetres and an actuating force of 0.30-0.35 N, equivalent to application of a weight of 300-350 g. The response key configuration and finger assignment are shown in Figure 5; the subset of keys used in each task, with an indication of the response corresponding to each, appear in Table 1. To avoid confusion, the keys should be labelled as appropriate for the task currently being performed.

If separate response keys cannot be interfaced to the computer, subjects' responses may be entered using the computer keyboard, substituting keys W, D, J, and I for response panel keys A, B, C, and D, respectively. This alternative arrangement should be adopted only if absolutely necessary, and should be recorded with the experimental data.

In the tracking task, the subject moves the joystick left or right to control the movement of a cursor on the screen of the computer monitor. The joystick lever and potentiometer should satisfy the following requirements:

i) The range of movement of the lever should be 30 degrees left and right from the vertical position.

ii) The friction of the moving parts should not exceed 50 g, and should be constant over the range of travel.

iii) The relationship between angular rotation of the joystick and lateral movement of the cursor should be linear for the entire range of travel.

iv) Analogue-to-digital conversion of joystick potentiometer values should be conducted to at least 8-bit resolution. In other words, rotation of the joystick should produce at least 256 discrete values.

Testing environment

External disturbances should be minimized during administration of the battery. If subjects are tested in groups, the test room should ideally be partitioned into separate workstations.

The position of the computer monitor relative to windows and sources of artificial light should be selected carefully, to avoid reflections on the screen. The surface of the screen should be perpendicular to the subject's line of sight, and located 0.6 metre from the eye; smaller or greater distances are acceptable if the size of individual characters is adjusted to maintain the visual angle within the specified range. The seat height should be about 0.45 metre, and the height of the upper surface of the response console about 0.75 metre.

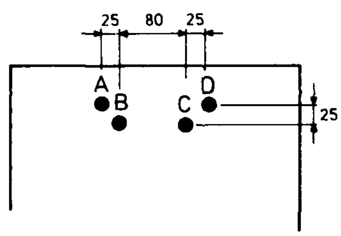


Figure 5. Response key configuration (all distances in millimetres). Key A is operated by the middle finger of the left hand; key B by the index finger of the left hand; key C by the index finger of the right hand; and key D by the middle finger of the right hand.

Table 1. Response keys used in each task. Key letter codes correspond to those in Figure 5.

TEST	KEYS USED	KEY ASSIGNMENT
REACTION TIME	A—D	Varies with condition
MATHEMATICAL PROCESSING	C, D or A, B	Right-handed subjects: C: < D: > Left-handed subjects: A: > B: <
MEMORY SEARCH	A, B or C, D	Right-handed subjects: A: YES B: NO Left-handed subjects: C: NO D: YES
SPATIAL PROCESSING	C, D or A, B	Right-handed subjects: C: SAME D: DIFFERENT Left-handed subjects: A: DIFFERENT B: SAME
UNSTABLE TRACKING	—	—
GRAMMATICAL REASONING	C, D or A, B	Right-handed subjects: C: SAME D: DIFFERENT Left-handed subjects: A: DIFFERENT B: SAME
DUAL-TASK	A, B or C, D	Right-handed subjects: A: YES B: NO Left-handed subjects C: NO D: YES

Training

Performance changes significantly as a task is learned. To avoid confounding between the effects of stressors and of task learning, the latter must be minimized or at least controlled. Ideally, subjects should practise the task until their performance is stable. The tasks comprising the STRES battery differ in the amount of practice necessary to achieve stability, and the mimimum requirements for each are specified in the task descriptions. If it is impossible to meet these requirements, an abridged practice schedule must be used to familiarize subjects with the task. Under these circumstances, particular attention must be given to inclusion of a suitable control group that is not exposed to the stressor but is otherwise tested under conditions identical to those of the experimental group. Moreover, if a within-subjects design is used in which each subject acts as his own control, the order of control and experimental conditions must be carefully balanced. Since the standard practice schedule is likely to produce more satisfactory results, the abridged schedule should be used only if absolutely necessary.

Task duration

The total duration of each task during the experimental (post-practice) phase is shown in Table 2, together with a summary of the amount of practice required. It is desirable to adhere to the duration specified for each trial block. However, if the effects of a stressor are unlikely to become apparent within this limited time period, a multiple of the specified value may be used.

Table 2. Summary of duration of each task during experimental sessions, and amount of practice required.

Task	Total duration of experimental test session (minutes)	Standard practice schedule (blocks)	Abridged practice schedule (blocks)
Reaction Time	15	Basic: 16 Other conditions: 4 each	Basic: 4 Other conditions: 4 each
Mathematical Processing	4	10	2
Memory Search	8	10 for each memory set size	2 for each memory set size
Spatial Processing	4	10	2
Unstable tracking	4	10	2
Grammatical Reasoning	4	8	2
Dual-task	8	5 for each memory set size	2 for each memory set size

F. TASK SPECIFICATIONS

REACTION TIME TASK

Purpose

The purpose of the Reaction Time (RT) task is to test the separate stages that comprise the reaction process. Basic RT is measured first, followed by four blocks of more complex trials, each loading a specific stage of the reaction process. The RT differences between complicated and basic blocks give specific information about the effect of loading four specific stages.

General Description

Digits are presented on a computer monitor, one at the time. The subject reacts to each digit by pressing the appropriate key on the response panel. S-R mapping is based on a) position of the digit, either left or right, and b) identity of the digit. Manipulated across trial blocks are the following task variables: *stimulus quality*, *compatibility of S-R mapping*, *time uncertainty* about stimulus onset, and response complexity.

Background

The idea that the process between stimulus presentation and overt reaction contains a number of discrete steps or stages is an old one. The first experimental studies on the duration of mental processing stages are attributed to Donders (1868), who tried to estimate the duration of decision processes by subtracting simple (non-choice) reaction times from choice reaction times. Donders' work was at least partly stimulated by a) Muller's incorrect pronouncement that nerve transmission time was "infinitely short" and could not be measured, b) Helmholtz's subsequent measurement of nerve conduction velocities and c) Hirsch's work on simple reaction times (Massaro, 1975). At the turn of the century Kulpe and co-workers criticized the subtractive method on the basis of introspective reports that it affected the 'Gestalt' of the tasks. Interest in Donders' method then waned, and was revived 100 years after it was first reported. Significant events were Posner and Mitchell's analysis of stimulus matching times in 1967, and initiation of the Attention and Performance symposia. The first three symposia were held in the Netherlands in the late sixties. The second, called the Donders Centenary Symposium on Reaction Time, contained contributions by Posner, Sanders, Sternberg, Welford, and many others. Especially important was Sternberg's 'Extensions of Donders' Method' (Sternberg, 1969b), which introduced the Additive Factor Method. The new method was based on the premise that processing stages can be identified by investigating the relation between different task variables rather than between different tasks as proposed by Donders.

The Additive Factor Method became an influential research method, and many studies on the effects of task variables were conducted. At least five different stages, or groups of stages, were identified, associated with (a) stimulus processing or encoding, (b) response choice, (c) motor programming, (d) motor activation, and (e) response execution. Based on these results, the following four task variables were selected for the current RT task: stimulus quality, compatibility of stimulus-to-response mapping, time uncertainty concerning stimulus onset, and response complexity. Figure 6 illustrates how these variables are assumed to map onto processing stages.

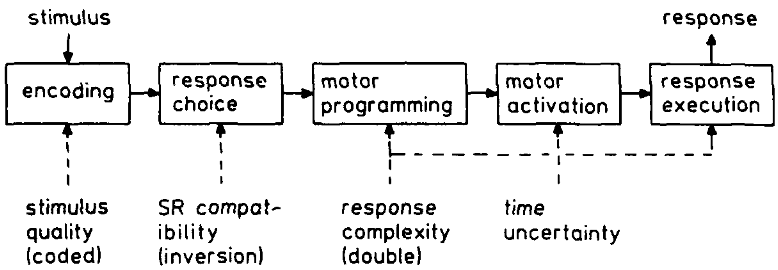


Figure 6. Stages of the reaction process, and the effects of some task variables.

Reliability

Split-half reliabilities for the Reaction Time task were computed by comparing scores in the first and second two minutes of a four-minute block. Data were obtained from a group of 158 subjects, aged between 18 and 19, of whom 14 were female. Reliability of mean RT for the Time Uncertainty block was 0.81, probably because the slow and irregular stimulus presentation decreased the number of trials completed; for the other blocks, it lay between 0.88 and 0.92. Error percentages were less reliable (0.32 for Time Uncertainty; 0.61-0.73 for the others).

More important are the split-half reliabilities of the difference scores corresponding to specific stages of the reaction time process. Reliabilities of these differences were between 0.62 and 0.74; reliability of response execution time was 0.94.

Validity

The question of validity is concerned primarily with the adequacy of the Additive Factor Method. The rationale of the Additive Factor Method is that two task variables are inferred to affect separate processing stages if they have additive main effects on RT, that is, if the size of the effect of one variable does not depend on the level of the other; and are inferred to affect at least one common processing stage if they have interactive effects on RT, that is, if the size of the effect of one variable does depend on the level of the other.

The four variables of the current task were tested in 2x2 factorial combinations as a final check on additivity. As shown in Figure 7, no interactions were obtained, supporting the claim that each variable affects a separate stage. Response execution time (not shown in the figure) was 552 milliseconds.

Sensitivity

The RT task has been shown to have non-specific sensitivity to factors such as fatigue and sleep loss, old age, brain damage, and a variety of drugs including barbiturates, amphetamines, and antihistamines (eg Boer, Ruzius, Mimpen, Bles, & Janssen, 1984; Gaillard, Gruisen, & de Jong, 1986; Gaillard, Rozendaal, & Varey, 1983; Gaillard, Varey, & Ruzius, 1985; Gaillard & Verduin, 1983; Moraal, 1982). Effects related to particular stages of the reaction process are somewhat more rare, but have been reported for the encoding stage by Logsdon, Hochhaus, Williams, Rundell, and Maxwell (1984), by Sanders, Wijnen, and van Arkel (1982), and by Steyvers (1987) with regard to sleep deprivation; by Frowein, Gaillard, and Varey (1981) with regard to a barbiturate; and by Stokx and Gaillard (1986) with regard to brain damage. Effects specifically related to the response-choice stage have been reported by Sanders et al (1982) for sleep deprivation; and by Stokx and Gaillard (1986) for brain damage. Effects specifically related to the motor-activation stage have been reported by Frowein, Reitsma, and Aquarius (1981) for sleep deprivation, and by Stokx and Gaillard (1986) with regard to brain damage. Specific effects on response execution have been reported by Frowein (1981) and Frowein et al (1981) for an amphetamine.

Technical Specification

A flow diagram of the structure of the task appears in Figure 8. The stimuli are shown in Figure 9. The subject places index and middle fingers of both hands on the response keys, as indicated in Figure 10. The response required to each stimulus is shown in Figure 11. Note that digits appearing on the left side of the screen require left-hand reactions, and that those appearing on the right side require right-hand reactions; this arrangement constitutes compatible S-R mapping. The distance between the left and right stimulus positions is 63 millimetres centre to centre; the size of the individual stimulus is 57 x 46 millimetres, including the rectangular frame.

Each trial has the following structure: 1) the stimulus is presented for one second; 2) the screen blanks for one second; 3) if the subject responds incorrectly within the first second, a feedback message (comprising the word 'error' or its equivalent) is presented immediately after the one-second stimulus presentation period; if the subject responds incorrectly during the blank interval, or fails to respond within two seconds of presentation of the stimulus, the feedback message is immediately presented for 0.5 second. The interval between presentation of successive stimuli is always at least one second. Trial duration is therefore two seconds if the subject responds correctly during this period, but may be lengthened by up to 0.5 second if an incorrect response, or no response, is made (see Figure 12). In each trial block, the stimulus is equally likely to be 2, 3, 4, or 5, and is equally likely to appear on the left or right.

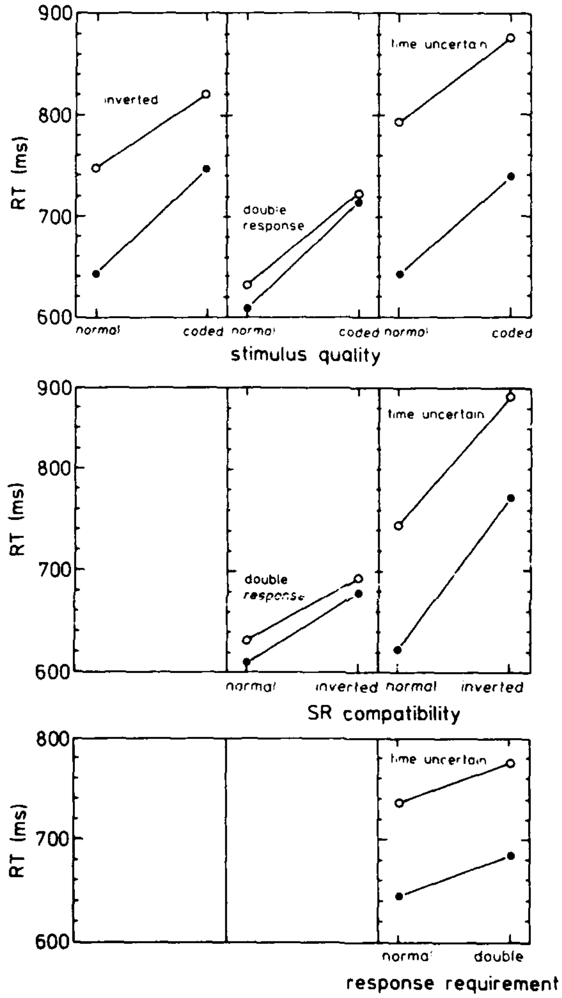


Figure 7. Experimental checks on the additivity of task variables. RT is shown as a function of stimulus quality (top panel), S-R compatibility (middle panel), and response complexity (bottom panel). Upper lines (open circles) refer to more complex conditions (inverted, double responses, time uncertain); lower lines (filled circles) refer to simpler conditions (noninverted, single response, and time certainty).

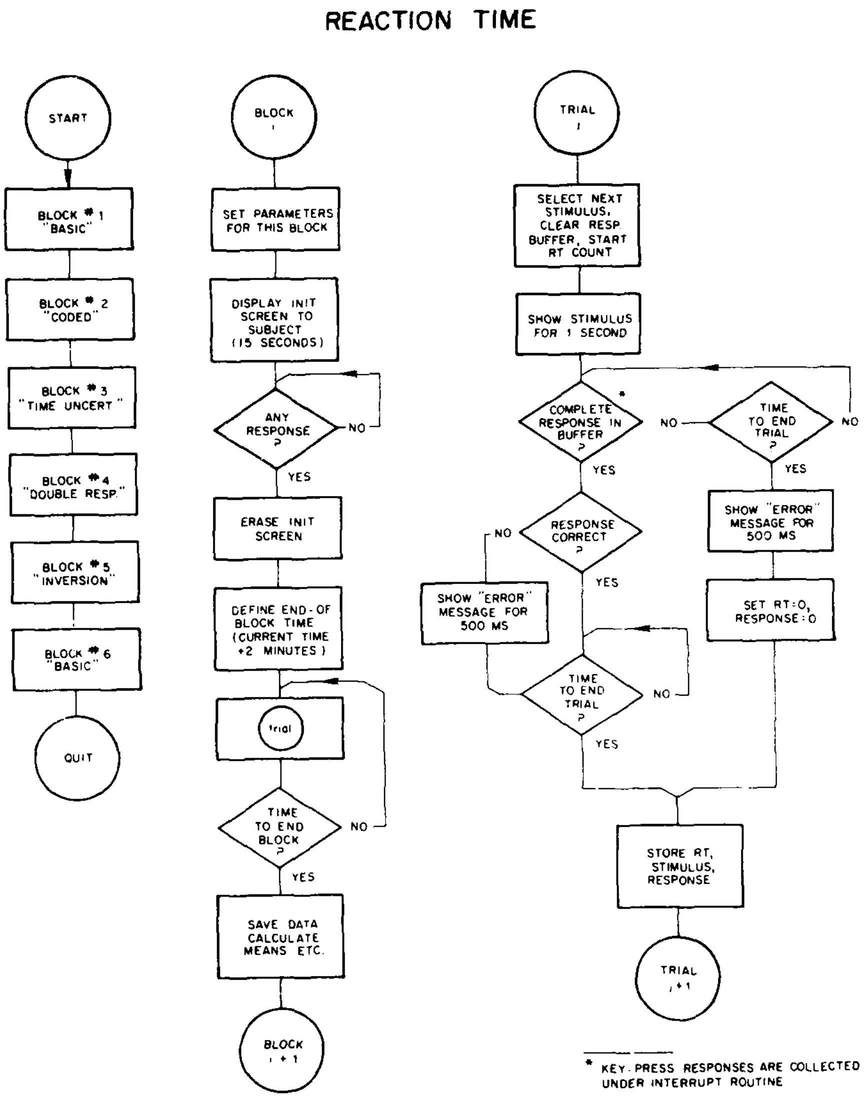


Figure 8. The structure of the Reaction Time task.

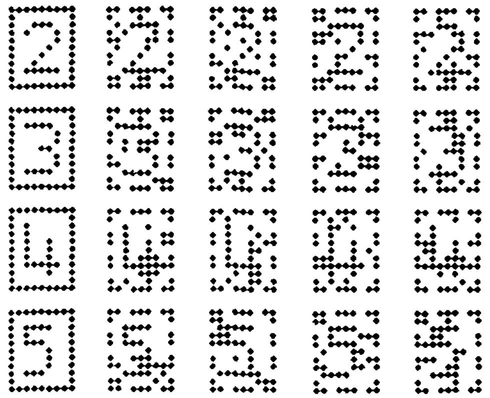


Figure 9. Normal and degraded stimuli used in the Reaction Time task. Stimuli are surrounded by a rectangular frame, measuring 57 x 46 millimetres. Four degraded versions of each digit are created by moving 10 elements from the frame towards the figure; the stimuli used in the task should be exactly as *illustrated*. Each element comprises two triangles, situated side by side with one pointing to the left and the other pointing the right to form a diamond shape. The grid on which the triangles are placed is the same as that used for normal presentation of text.

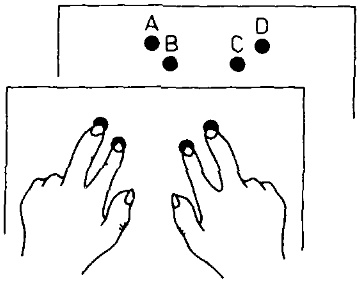


Figure 10. Assignment of fingers to response keys in the Reaction Time task.

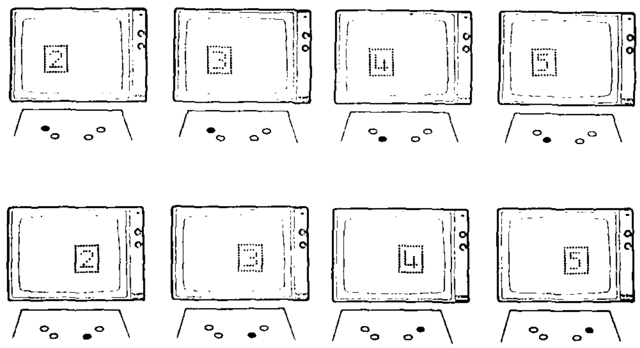


Figure 11. Stimuli and responses in the Reaction Time task. The response key required by the stimulus is indicated in black.

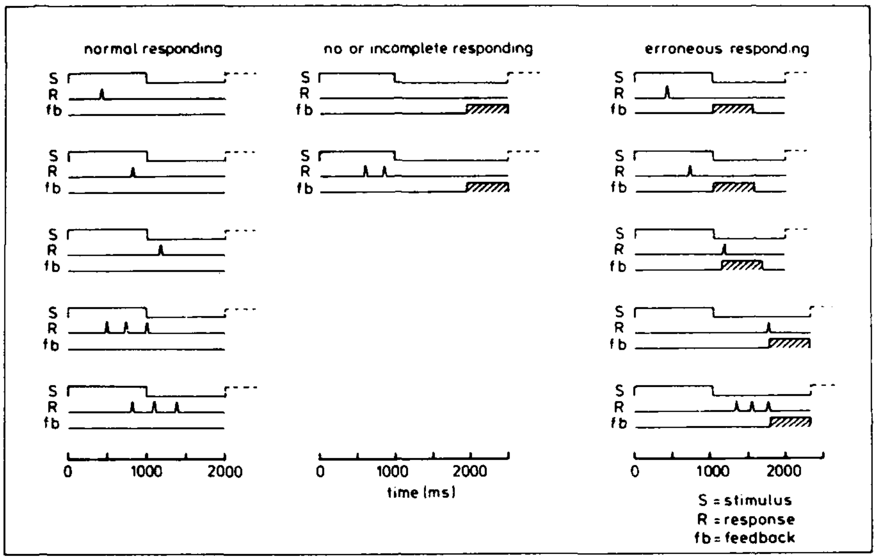


Figure 12. Stimulus, response, and feedback: the three components of a trial in the Reaction Time task.

Incomplete responding may occur in the Double Responses block, where each stimulus requires a predefined sequence of three key-presses. Failure to complete the sequence before the blank period expires triggers the feedback message (Figure 12).

Trial blocks are of two-minute duration, and usually comprise 60 trials. Each block begins with a brief announcement reminding the subject of the nature of the experimental condition. After 15 seconds, a flashing message instructs the subject to initiate the block by pressing one of the response keys. The first seven trials are for 'warming up', and are excluded from subsequent analysis. The message 'end of block' is presented on the screen after the final response.

The standard procedure consists of an instruction phase of about five minutes, followed by the practice phase, a break of at least five minutes, and 15 minutes of data collection. Experimental data collection is conducted in four complex blocks (eg 'Coded') preceded and followed by a Basic block, yielding a total of six blocks.

The blocks are administered in the following order:

1. Basic: The S-R mapping is as shown in Figure 11. Stimulus quality is normal, and the inter-stimulus interval varies from two seconds (for a correct response) to a maximum of 2.5 seconds (for an incorrect response or a response failure).

2. Coded: Identical to Basic, except that stimulus quality is low. Each of the four degraded versions of each digit is equally likely to be presented.

3. Time Uncertainty: Identical to Basic, except that a) stimuli are presented irregularly by means of variable interstimulus intervals (ISIs) chosen randomly to assume any integer value between 2000 and 10000 milliseconds, and hence b) there are approximately 22, rather than 60, stimuli.

4. Double Responses. Identical to Basic, except that, instead of a single key-pre..s for each stimulus, three keys must be pressed in a particular order. For example, a 2 on the left side of the screen, normally requiring a single key-press with the left middle finger, now requires the following sequence of key-presses: left middle, left index, left middle. Thus, the normal A response is replaced by the ABA sequence; BAB replaces the B response; CDC the C response; and DCD the D response. RT is defined as the interval between stimulus onset and first key-press response; response execution time is the interval between first and last key-press.

5. Inversion. Identical to Basic, except that the S-R mapping is made incompatible by requiring a left-hand key-press response for stimuli on the right side of the screen, and a right-hand key-press response for stimuli on the left side of the screen. For example, a 2 on the left side of the screen requires depression of key C by the right index finger.

6. Basic (during data collection phase only).

Data Specification

The RT for each trial is coded as positive for a correct response, negative for an incorrect response, and 0 for a response failure. Recorded for every trial are 1) a stimulus-code (digit identity, position, and quality); 2) a response code (key identity); and 3) RT.

For each block, the following summary statistics are calculated: a) mean RT for correct responses; b) the standard deviation (SD) of RTs for correct responses; c) number of trials; d) percent errors (excluding response failures); and e) percent response failures.

Normative Data

Normative data have been collected for 450 subjects, including 26 females, aged between 16 and 32 years (mean = 21.6). The standard procedure was followed except for some details, the most notable of which was the use of four-minute rather than two-minute blocks.

Mean RT in the first and second basic block was considered a nonspecific component (or remnant) of the reaction process; the differences between mean basic RT and mean RT in each of the four complex blocks were considered to represent measures of four specific stages of the reaction process.

As shown in Table 3, five performance evaluation categories were defined. These were based on frequency distributions of the individual subjects. Categories were tentatively labelled as 'very good', 'good', 'average', 'poor', and 'very poor'. Each category is based on the range of performance achieved by 20% of the subjects. Thus 'good', for example, represents the 90 subjects falling within the 60th-80th percentile range. The table also gives percentage error, unless reliability was below 0.40.

Table 3. Normative data for the RT task (n=450). Evaluation categories are based on frequency distributions. 'Very good' is the performance level of the best 20% of the subjects, 'good' the performance level of the next 20%, and so on. All scores except Basic RT are difference scores between complex and basic blocks.

	evaluation category					
	v. good	good	average	poor	v. poor	split-half reliability
RT in milliseconds						
Basic RT	<566	566-602	603-642	643-692	>692	0.95
Encoding	< 65	65- 84	85-105	106-139	>139	0.67
Resp. Choice	< 60	60- 86	87-107	108-134	>134	0.74
Motor Prog.			< 16	17- 41	>41	0.65
Motor Act.	< 57	57- 86	87-113	114-139	>139	0.62
Resp. Exec.	<445	445-507	508-557	558-630	>630	0.94
Percent error						
Basic RT	< 1.0	1.0-1.8	1.9-3.2	3.3-5.4	>5.4	0.78
Encoding	< 0.8	0.9-2.4	2.5-5.4	5.4-11.5	>11.5	0.49
Resp. Choice	< 1.8	1.9-3.8	3.9-6.4	6.5-9.5	>9.5	0.52
Motor Prog.				<= 0.5	>0.5	0.43

Training Requirements

Some studies have investigated the effects of extensive training on the Reaction Time task. Boer (1987) tested 32 subjects using blocks of 24 minutes on separate days. Blocks contained a random mixture of normal and low-quality stimuli. RT during the first eight minutes decreased over blocks. Relative to the initial level of Block 1, decrements of 11%, 16%, and 19% were observed during the initial periods of Blocks 2, 3, and 4, respectively, suggesting that performance may reach a stable level after some 2000 trials.

Fewer training data are available for specific effects. There is a clear suggestion that training reduces the effect of stimulus quality. For example, when 12 subjects completed two 640-trial sessions on four successive days, the degradation effect was reduced from 107 milliseconds in the first session to 85 milliseconds in the last.

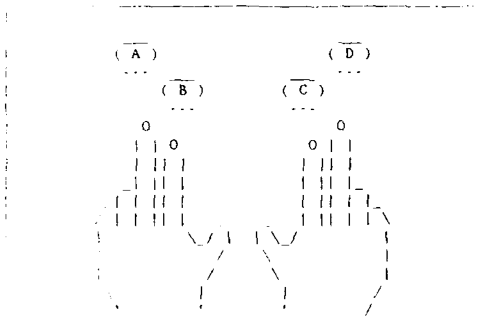
The standard training schedule for this task comprises 16 blocks of the Basic condition, followed by four blocks of each of the remaining conditions. The abridged schedule comprises administration of four Basic blocks followed by one of each remaining condition. No data are collected during training.

Instructions to Subjects

This is a test of the speed and efficiency of your reactions. You should respond as quickly as possible, but avoid errors. Slow down a little if you start to make errors, because this probably means that you are going beyond your capacity, but don't be too concerned about an occasional error.

After you have read these instructions, you will be given the opportunity to practise the task. This will be followed by about 15 minutes of actual measurement of your performance.

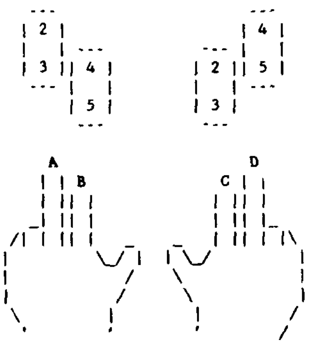
Before the task begins, you should place your fingers on the response keys as illustrated in the diagram below, and respond to each signal or stimulus appearing on the computer monitor by pressing one of the keys.



The diagram indicates how you should place the fingers of your left hand on buttons A and B, and the fingers of your right hand on buttons C and D. Keep your fingers there as long as the task runs--reactions are faster that way.

The diagram indicates how you should place the fingers of your left hand on buttons A and B, and the fingers of your right hand on buttons C and D. Keep your fingers there as long as the task runs -- reactions are faster that way.

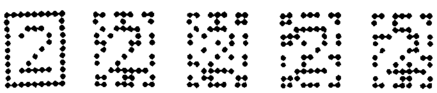
If a signal appears on the left side of the screen, use your left hand. If a signal appears on the right side, use your right hand. So, digit position (left/right) immediately tells you which hand to use. The signals are the digits 2-5. Use button A or C for 'low' digits (2 and 3), and button B or D for 'high' digits (4 and 5). The combination of hands and fingers is as follows: Suppose you get a '3' on the left side of the screen. Left side means left hand, so it can only be button A or B. Digit 3 is low, so that means it has to be button A. Another example: If you get a '4' on the left side, you react with button B. The diagram below illustrates the rules of the task.



In each measurement block, you should respond by pressing the appropriate key each time a signal appears. Keep your fingers on the response panel throughout each block, and relax during the short breaks between blocks.

The blocks are:

1. Basic block. You already know what to do in a basic block.
2. Coded block. Same as Basic, but the digits are more difficult to see. The diagram shows some examples.



3. Time Uncertainty. Same as Basic, but the digits come at irregular times, and sometimes unexpectedly.

4. Double Reaction. Same as Basic, but you press three buttons for every digit. The rule is this: The first button is the regular one, as in Basic; the second is the other one on the same hand; the last is the same as the first. For example, suppose you get a '4' on the right side. The regular button is D. So now you press DCD in that order. To sum up: ABA instead of just A; BAB instead of B; CDC instead of C; and DCD instead of D.

5. Inversion. Use the left hand if the digit is on the right side and use the right hand if the digit is on the left side. The rule left side — left hand; right side — right hand is now reversed.

6. Basic. The last block is Basic again.

Before each block begins, you will be reminded what you have to do. Place your fingers on the response keys, press one of the keys, and the block will begin. Go as fast as possible, but mind the errors.

Please press any response key to proceed.

[Instructions given immediately prior to a Basic block:]

THIS IS A BASIC BLOCK. Remember:

If a signal appears on the left side of the screen, use your left hand. If a signal appears on the right side, use your right hand. So, digit position (left/right) immediately tells you which hand to use. The signals are the digits 2-5. Use key A or C for 'low' digits (2 and 3), and key B or D for 'high' digits (4 and 5). The combination of hands and fingers is as follows: Suppose you get a '3' on the left side of the screen. Left side means left hand, so it can only be key A or B. Digit 3 is low, so that means it has to be key A. Another example: If you get a '4' on the left side, you press key B.

Please place your fingers on the keys, and press any key to begin the block.

[Instructions given immediately prior to a Coded block:]

THIS IS A CODED BLOCK. Remember:

In this block, the digits are more difficult to identify, but the task is otherwise the same as in the Basic block. So if the signal appears on the left, use your left hand; if it appears on the right, use your right hand. Use key A or C for 'low' digits (2 and 3), and key B or D for 'high' digits (4 and 5).

Please place your fingers on the keys, and press any key to begin the block.

[Instructions given immediately prior to a Time Uncertainty block:]

THIS IS A TIME UNCERTAINTY BLOCK. Remember:

In this block, the digits are presented at irregular intervals, but the task is otherwise the same as in the Basic block. So if the signal appears on the left, use your left hand; if it appears on the right, use your right hand. Use key A or C for 'low' digits (2 and 3), and key B or D for 'high' digits (4 and 5).

Please place your fingers on the keys, and press any key to begin the block.

[Instructions given immediately prior to a Double Reaction block:]

THIS IS A DOUBLE REACTION BLOCK. Remember:

In this block, you press three keys for every digit (ABA instead of A, BAB for B, CDC for C, and DCD for D), but the task is otherwise the same as in the Basic block. So if the signal appears on the left, use your left hand; if it appears on the right, use your right hand. Press ABA or CDC for 'low' digits (2 and 3), and BAB or DCD for 'high' digits (4 and 5).

Please place your fingers on the keys, and press any key to begin the block.

[Instructions given immediately prior to an Inversion block:]

THIS IS AN INVERSION BLOCK. Remember:

In this block, you use your left hand if the digit appears on the right, and your right hand if it appears on the left, but the task is otherwise the same as in the Basic block. So press key A or C for 'low' digits (2 and 3), and key B or D for 'high' digits (4 and 5).

Please place your fingers on the keys, and press any key to begin the block.

MATHEMATICAL PROCESSING TASK

Purpose

The purpose of this mental arithmetic task is to place demands upon processing resources associated with working memory. Specifically, the subject is required a) to retrieve information from long term memory, b) to update information in working memory, c) to execute arithmetical operations sequentially, and d) to perform numerical comparisons.

General Description

This test requires subjects to perform two arithmetical operations, addition and/or subtraction, on a set of three single-digit numbers, and to determine whether the answer is greater than or less than five. Problems are presented in the centre of the monitor screen in a horizontal format (eg $5 + 3 - 4 =$); *the subject is instructed to solve the problem working from left to right*, and to press the key marked '>' or '<'. The duration of each trial block is three minutes. On each trial, RT is recorded from onset of the problem to execution of a response.

Background

The present test, developed by Shingledecker (1984), requires the execution of two mathematical operations (addition and/or subtraction) within a given problem. In this section, the literature on mathematical processing is reviewed briefly.

Chiles, Alluisi, and Adams (1968) developed a mathematical processing task, requiring both addition and subtraction, for use in the assessment of mental workload. This task was included in the Multiple Task Performance Battery (MTPB) with other cognitive tasks such as auditory vigilance, warning lights, meter monitoring, problem solving, choice reaction time, tracking, and pattern discrimination; it was used in multi-task studies to examine subjects' time-sharing ability (eg Chiles & Alluisi, 1979; Chiles, Bruni, & Lewis, 1969; Chiles & Jennings, 1970; Hall, Passey, & Meighan, 1965).

Perez (1982) examined working memory storage and processing in the solution of multi-operation problems. RT and accuracy for problems involving three operations (combinations of addition and subtraction) were examined in five experiments. The arithmetical notation (eg algebraic or reverse Polish) was varied to investigate subjects' ability to manipulate arithmetical information. The results showed that a) errors in computation were a function of loss of operand information (the digits) and confusion between operations (eg adding instead of subtracting); b) RT was a function of the number of different operations in a problem (e.g., +-+ was slower than +++); and c) after very little practice with the unfamiliar reverse Polish notation, which minimizes transient memory load, performance was superior to that obtained with algebraic notation.

Wanner and Shiner (1976) also employed multi-operation problems in the study of working memory. Their experiment focused on the transient memory load imposed by problems involving two operations of subtraction, with parentheses appearing either on the left, as in (5-4)-1, or on the right, as in 5-(4-1). Each problem appeared sequentially, from left to right, and was interrupted at various points by presentation of a series of words; subjects were then required, with equal probability, to solve the problem or recall the words. Wanner and Shiner found that errors on the word-memory task and the mathematical task were related to the transient memory load imposed by pending operations. For example, the transient memory load for the right-parentheses problems is greater than that for the left-parentheses problems, since subjects must defer computation until the entire problem has been presented.

Finally, Shingledecker (1984) used multi-operation mathematical reasoning problems in the development of a standardized loading task. Three distinct levels of task demand were selected empirically on the basis of RT and accuracy data obtained for factorial combinations of total number of operations and sequence of addition and subtraction operations within the problem. The version of the task used in the STRES Battery corresponds to the moderate demand level identified by Shingledecker.

The Mathematical Processing task is assumed to tap primarily central processing resources ('higher mental processes'); its demands on input and output stages are minimal. Performance on the task may be broken down into four processing stages: a) retrieval of arithmetical information from long term memory, b) updating of information in working memory, c) sequential execution of arithmetical operations, and d) numerical comparison. These processes are considered in more detail below.

Ashcraft and Battaglia (1978), Ashcraft and Stazyk (1981), and Stazyk, Ashcraft, and Hamann (1982) have investigated the role of retrieval from long term memory in the solution of simple arithmetical problems by adults. It appears that adults rely on a well organized memory structure rather than procedures such as counting; in effect, 'mathematical tables' are stored in their long term memory.

Problems involving multiple operations require subjects to carry out different arithmetical operations rapidly and sequentially. They must also maintain and update a sequence of sub-totals. For example, the problem '$7 + 2 - 3 + 1 - 4$' produces the sequence 9, 6, 7, 3. This type of activity requires both storage (eg Wanner & Shiner, 1976) and processing in working memory. Previous research (eg Perez, 1982) has shown that transitions from one operation to another (eg +,-) require more time than sequential use of the same operator (eg +,+), perhaps because of a memory priming effect for arithmetical operations.

The processes involved in comparison of an internally generated answer against a standard value were investigated by Restle (1970), who required subjects to compare the sum of two numbers (A + B) to a standard (C) and select the greater of the two. Response latency was inversely related to the numerical difference between the sum and the standard, suggesting an analogue operation in which the magnitudes (A + B) and C were mapped onto an internal number line prior to comparison.

Reliability

Reliability information for the STRES version of the Mathematical Processing task has been provided by Schlegel and Gilliland (in press). They reported a reliability coefficient of 0.85 for a group of 123 subjects who had practised the task for five three-minute blocks of trials, one block per day. The reliability was estimated between data collected on two separate days after the five practice blocks, with one day separating the two tests.

In addition, reliability data have been obtained in a paper and pencil arithmetic test involving addition or subtraction of two three-digit numbers, multiplication of two two-digit numbers, and division of a four-digit number by a two-digit number (Seales, Kennedy, & Bittner, 1980). Eighteen subjects were tested on 15 consecutive days, completing 64 problems per day during the first seven days and 96 problems per day thereafter. Performance (total number of problems attempted, total correct, and correct minus wrong) showed improvement over the first nine days of testing and remained stable thereafter. In addition, the inter-day correlations for the above three measures were relatively high (mean $r = 0.935$, 0.941, and 0.921, respectively).

Chiles, Jennings, and Alluisi (1978) reported reliability coefficients for a multi-operation task requiring the addition of two two-digit numbers and the subtraction of a third two-digit number (eg 12 + 15 − 13 =). There were 94 subjects in this study, but only 51 were tested on two consecutive days. Subjects received 15 minutes of practice before the start of testing. The arithmetic task was performed in conjunction with a problem solving, manual tracking, or monitoring task. The authors computed reliability coefficients by correlating performance on the mathematical task across all task combinations. The average correlations for those subjects tested for one day were 0.73 and 0.82 for solution time and accuracy, respectively; for those subjects tested on two consecutive days, the average correlations were 0.91 and 0.71 for solution time and accuracy, respectively.

Validity

As discussed above, research with single-digit addition problems (eg Ashcraft & Stazyk, 1981) has supported the hypothesis that adults solve simple addition problems by recourse to information stored in long term memory. Moreover, research with multi-digit addition problems (eg Hitch, 1978) has shown that complex mathematical problems are solved in a series of elementary steps requiring storage in working memory.

Chiles et al (1978), using multi-operation problems, reported a pattern of dual-task interference consistent with the notion that mathematical processing taps working memory resources. Performance on an arithmetic task was poorer with a concurrent code lock solving task than with a concurrent manual tracking task that placed demands primarily upon response-based resources.

Sensitivity

The STRES version of the Mathematical Processing task has been employed in the study of the effects of caffeine and 24 hours' sleep loss. Schlegel and Gilliland (in press) reported significant increases in RT in a study using two mg/kg and four mg/kg caffeine with three levels of difficulty of the Mathematical Processing task, including the level of difficulty specified for the STRES battery. They also found that sleep loss produced significant slowing of RT in this task over all three levels of difficulty, and for the specific level of difficulty used in the STRES battery.

Data are also available for tasks similar to that used in the STRES battery. Repko, Jones, Garcia, Schneider, Roseman, and Corum (1976), for example, reported an effect on mathematical processing of exposure to methyl chloride (35 parts per million).

The pattern of dual-task interference noted by Chiles et al (1978) suggests that mathematical processing is likely to be most sensitive to stressors that affect working memory. This conclusion is, however, tentative. More detailed evidence of sensitivity will emerge as the STRES data base accumulates.

Technical Specification

The structure of the task is illustrated in Figure 13. The duration of each trial block is three minutes. Problems are presented in the centre of the monitor screen, and comprise three operands (each a single digit) separated by two arithmetical operators (+ or -) and followed by =. Each character subtends 15-20 minutes of arc at a viewing distance of 0.6 metre. The operands and operators comprising each problem are randomly selected with the following constraints: 1) only the digits 1-9 are used; 2) the correct answer may be any number from 1 to 9 except 5; 3) the answers 'less than 5' and 'greater than 5' are equiprobable within a trial block; 4) cumulative intermediate totals, working from left to right, must have a positive value; 5) the same digit must not appear twice in the same problem, unless it is preceded by the same operator on each occasion (eg +3 and +3 is acceptable; +3 and -3 is not); and 6) the sum of the absolute value of the digits in a problem must be greater than 5. Example problems are shown in Table 4.

Table 4. Examples of problems in the Mathematical Processing task.

Problem	Correct response
$6-5+2=$	$<$
$9-1-2=$	$>$
$2+6-4=$	$<$

MATHEMATICAL PROCESSING

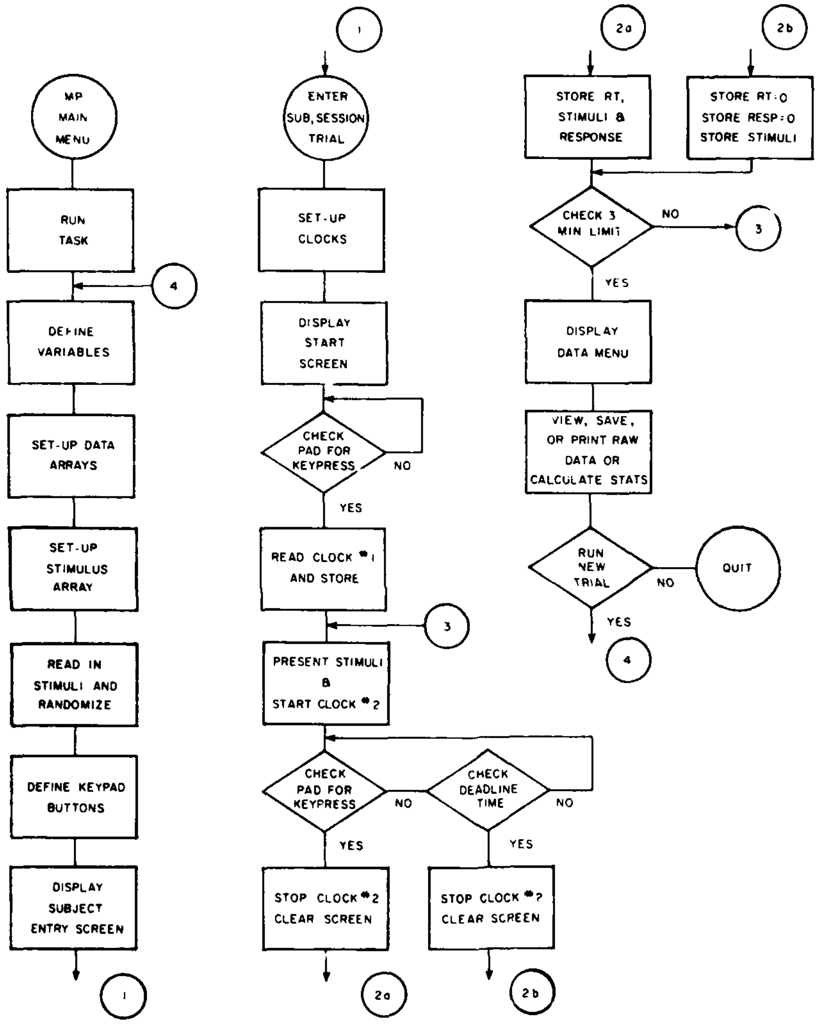


Figure 13. The structure of the Mathematical Processing task.

The subject responds to each problem by pressing one of two keys to indicate whether the answer is greater than (>) or less than (<) 5. A sample stimulus display is shown in Figure 14.

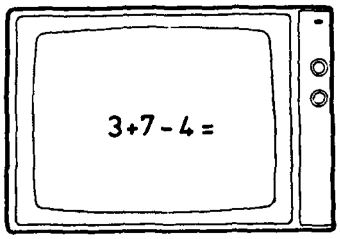


Figure 14. Example of the stimulus display in the Mathematical Processing task.

Experimental and practice trials have the following structure: 1) a problem is presented in the centre of the monitor screen; 2) as soon as the subject responds, or a deadline of 15 seconds has elapsed, the problem is erased; 3) the screen blanks for an interstimulus interval varying ran.: mly between 3000 and 5000 milliseconds; and 4) a new problem is presented.

Demonstration trials differ from the experimental and practice trials as follows: 1) as soon as a response is made, the problem is accompanied by an indication of the correct solution, the response made, and the RT (see Figure 15); 2) this feedback remains on the screen until the subject presses either response key to initiate the variable ISI, as in step (3) above.

After the final trial in any block, the message 'end of block' appears in the centre of the screen.

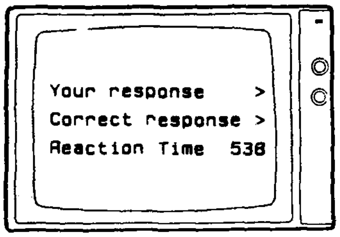


Figure 15. Example of feedback given during demonstration blocks in the Mathematical Processing task.

Data Specification

Each RT is coded as positive for a correct response, negative for an incorrect response, and 0 for a response failure. For every trial within a three-minute trial block, the following data are recorded: 1) composition of the problem, and 2) RT.

The following summary statistics are determined for each block: a) mean of all correct RTs; b) SD of all correct RTs; c) mean of correct RTs for response 'greater than'; d) SD of correct RTs for response 'greater than'; e) mean of correct RTs for response 'less than'; f) SD of correct RTs for response 'less than'; g) number of 'greater than' problems completed; h) number of 'less than' problems completed; i) percent errors to 'greater than' problems ; j) percent errors to 'less than' problems; k) percent response failures for 'greater than' problems; and l) percent response failures for 'less than' problems. In the calculation of error rates (i-j), response failures are excluded.

Training Requirements

Subjects are given the opportunity to read the instructions, and are then presented with 10 demonstration trials. During these trials, the experimenter should monitor the subject's performance to determine whether the instructions are being followed. In particular, it should be ensured that the problems are solved from left to right to avoid negative intermediate results, and that a suitable speed/accuracy compromise is maintained.

Three-minute practice blocks are then administered. The standard training schedule is 10 blocks, the abridged schedule is two blocks.

To summarize, the following procedure should be adopted:

1. Present instructions to the subject.

2. Run the demonstration trials, monitoring the subject's performance to ensure that the instructions are being followed.

3. Run the practice trial blocks.

Note that, if the task is administered to the subject in several sessions, the demonstration and practice trials should be omitted after the first session.

Instructions to Subjects

Demonstration trials:

In this task, you must solve a number of simple addition and subtraction problems to determine whether the correct answer is less or greater than 5. The two possible responses are 'less than' (<) and 'greater than' (>), and these are entered by pressing the appropriately labelled key on the response console.

Please start the task whenever you are ready by pressing either of the response keys. There are 10 demonstration problems in this block. The problems appear one at a time on the screen, and should be solved from left to right. Each problem requires two operations (addition and/or subtraction). Always perform the additions and subtractions in the order that they appear in the problems. As soon as you respond to a problem, you will be informed of your reaction time and accuracy. When you are ready to proceed to the next trial, press either of the response keys; the display will be erased and the next problem will appear shortly afterwards. Try to perform the task as quickly and accurately as possible. Go as fast as you can, but if you start to make errors because you are trying to go too fast, slow down. You should try to respond correctly to every problem. After you have completed the 10 demonstration trials, the message 'end of block' will appear.

Experimental and practice blocks:

In this task, you must solve a number of simple addition and subtraction problems to determine whether the correct answer is less or greater than 5. The two possible responses are 'less than' (<) and 'greater than' (>), and these are entered by pressing the appropriately labelled key on the response console.

Please start the task whenever you are ready by pressing either of the response keys. Testing periods last for three minutes each. The problems appear one at a time on the screen, and should be solved from left to right. Each problem requires two operations (addition and/or subtraction). Always perform the additions and subtractions in the order that they appear in the problems. As soon as you respond to a problem, it will be erased and a new problem will appear shortly afterwards. Try to perform the task as quickly and accurately as possible. Go as fast as you can, but if you start to make errors because you are trying to go too fast, slow down. You should try to respond correctly to every problem. At the end of the three-minute testing period, the message 'end of block' will appear.

MEMORY SEARCH TASK

Purpose

This task examines the ability to search items held in memory for the presence of a 'probe' item. It is based on information-processing principles and additive-factor methodology, and can be used to investigate the loci of stressor effects.

General Description

This task is based on the paradigm described by Sternberg (1966, 1967, 1969a, 1969b, 1971). A set of letters (the 'memory set') is presented on a video monitor, followed by a single letter (the 'probe letter'). The subject has to indicate, by pressing an appropriate key, whether the probe letter is a member of the memory set. For example, if the memory set were G, X, T, L and the probe letter were T, then the correct response would be 'yes'; if the probe letter were D, then the correct response would be 'no'. The number of letters in the memory set can be varied to affect the difficulty of the task, and the major dependent variable is RT.

There are three main variations on the basic procedure. The Varied Set procedure involves presentation of a different memory set, followed by a single probe item, on every trial. The Fixed Set procedure presents one memory set followed by many (eg 100) probe items. The Mixed Set procedure is a mixture of the two, such as ten separate memory sets each followed by ten probes. The Fixed Set procedure is used here to conform to the requirements of the tracking task when this and Memory Search are co-administered.

Background

To perform the memory search task correctly, the subject must carry out several operations in sequence. First, he must memorize the memory set. This process must be completed before presentation of the probe item, otherwise it will contaminate the RT (recognition and storage of digits or letters take typically 250-500 milliseconds per item). When the probe item is presented, the subject must first detect and recognize it. He must then perform some sort of search and comparison of the probe item with the items held in memory. The outcome of this process provides the necessary information for the subject to select an appropriate response. Thus, the task includes detection, recognition, memory search and comparison, and response selection stages.

Variation of memory set size does not affect detection or recognition of the probe, or selection of the response; however, it does affect the intervening memory search and comparison stage. Thus, changes in RT with changes in memory set size can be used to determine the nature of the memory search process. Two basic memory searching algorithms can be identified, which predict two different reaction time functions: a serial search, which can be either self-terminating or exhaustive, and a content-addressable search (Massaro, 1975).

In a serial search, the memory set items are stored in separate addresses in memory and the probe item is compared successively with the contents of each address. In a serial, self-terminating search, the search stops when a match is found, or continues to the end if a match is not found. The probe, if present, is equally likely to appear in any memory set position. Thus, when RT is plotted against set size, the slope of the function for 'yes' responses will be about half that for 'no' responses, since only half of the memory set on average need be searched to find a match. In a serial exhaustive search, the search continues to the end whether or not a match is found. In this case, the functions for 'yes' and 'no' responses will have identical slopes.

In content-addressable search, memory locations are reserved for all items in the population from which the memory sets are drawn, and each is given the content 'no'. For example, if the items are digits, then 10 locations are labelled 0-9, and assigned the content 'no'. As each item in the memory set arrives, the content of its corresponding address is changed from 'no' to 'yes'. For example, if the memory set is 3, 7, 2, then the contents of addresses 3, 7 and 2 are changed to 'yes'. When the probe item arrives, its corresponding address is accessed and the answer is immediately available. In this case, changes in memory set size will not affect memory search time; in other words, the slope of the RT function will be zero for both 'yes' and 'no' responses.

It is probable that, in real life, search strategies vary with the information content of the memory set items (eg whether '4' is in the telephone number or whether 'butter' is in the refrigerator). Sternberg (1966) found that RT increased linearly as a function of memory set size, and that the 'yes' and 'no' functions had the same slope, indicating that his subjects had used a serial exhaustive search strategy. This conclusion was subsequently confirmed by many other investigators.

In another study, Sternberg (1967) covaried both the memory set size and the quality of the probe digit. On half of the trials, the probe digit was presented intact, and on the remaining trials it was degraded by placing it behind a masking screen of dots. A fixed-set procedure was used. Logically, it should take longer to recognize a degraded digit than an intact digit. Thus, the overall RT to degraded stimuli should be longer than that to intact stimuli. Further, it seems reasonable to assume that once the recognition stage has given the probe item a label, however easy or difficult it may have been to do so, the rate of memory search will be the same. Thus, the slope of the function should not change. If this is the case, the RT function for the degraded probe will have the same slope as that for the intact probe, but a higher intercept; if stimulus quality does affect memory searching time, however, then the 'degraded' slope will be greater than the 'intact' slope.

Sternberg found that degradation of the probe affected only the intercept of the RT function, indicating that this manipulation affected the recognition stage but not the memory search stage. Thus, it could be concluded that the probe was initially 'cleaned up' prior to memory search, increasing RT by a constant amount regardless of memory set size.

This rationale may be applied to other experimental variables. Generally, if task variables have additive main effects on reaction time, then they are inferred to affect separate processing stages. If they have interactive main effects, then they are inferred to affect at least one common processing stage. Thus, in the Sternberg test, an experimental variable that interacts with memory set size may be assumed to affect memory search, whereas a variable whose effect is additive to memory set size can be assumed to affect a stage other than memory search.

Methodological Variations

Many variations on Sternberg's original method have been studied, and reviews have been published by Hann (1973) and by Sternberg (1975). The main findings are summarized below in seven groups identified by Hann.

1. Stimulus Category and Quality.

Formally, or physically, similar stimuli are scanned more rapidly than stimuli with only associational similarity. Also, stimuli in the same modality are scanned more rapidly than those in different modalities (Lively & Sanford, 1972; Klatzky, Juola, & Atkinson, 1971; Naus, Glucksberg, & Ornstein, 1972).

2. Stimulus Probability and Frequency.

RT is inversely related to the probability of occurrence of a particular item belonging to the memory set, whether the item is repeated, specifically cued, or simply occurs more often over a series of trials (Briggs & Swanson, 1969; Theios, Smith, Haviland, Traupmann, & Moy, 1973).

3. Temporal Variables.

Varying the presentation rate of the memory set items has little or no effect on RT (Burrows & Okada, 1971), but changing the delay between the memory set and the probe item affects processing of the memory set. At short delays, memory search and comparison are held up until memory set processing is complete (Connor, 1972).

4. Spatial and Numerical Separation.

RT is faster when the stimuli are organized, such as in numerical sequence, and is faster on negative trials as a function of the numerical separation between the probe and the memory set (Morin, DeRosa, & Stultz, 1967; DeRosa & Morin, 1970).

5. Instructional Variables.

Emphasis on speed or on accuracy each produces strong practice effects on the intercept, but not the slope, of the RT function (Lively, 1972). RT is decreased with increasing delay of a probe after presentation of items which the subject has been told to remove mentally from the memory set (DeRosa, 1969; DeRosa & Sabol, 1973).

6. Probe Set Size

RT decreases as a function of the number of items common to the memory and probe sets (Briggs & Blaha, 1969; Briggs & Swanson, 1969; Briggs & Johnsen, 1973).

7. Miscellaneous Variables.

RTs to pictorial stimuli are faster when processed by the right cerebral hemisphere, and RTs to letters are faster when processed by the left hemisphere. When stimuli are presented to the 'slow' hemisphere for that type, the intercept of the RT function increases but the comparison rate is unaffected (Klatzky & Atkinson, 1971).

Linear and increasing RT functions have been observed for a wide variety of stimuli, including visual and auditory digits and letters, two-and three-digit numbers, shapes, pictures of faces, drawings of common objects, words of various lengths, colours, and phonemes (Burrows & Okada, 1973; Chase & Calfee, 1969; Clifton & Tash, 1973; Foss & Dowell, 1971; Hoving, Morin, & Konick, 1970; Orenstein & Hamilton, 1977; Swanson, Johnsen, & Briggs, 1972). The slopes of the RT functions to these types of stimuli differ systematically. The 'yes' and 'no' functions have been found to remain linear and parallel for memory sets of up to ten letters (Wingfield & Branca, 1970) and up to twelve common words (Naus, 1974).

Individual Differences

Linear and increasing RT functions have been observed in people of differing personalities, various ages ranging from children to elderly adults, and in normals, alcoholics, schizophrenics, and the brain-damaged mentally retarded. Aging and mental retardation both produce increased slopes compared with young, healthy adults (Anders, Fozard, & Lillyquist, 1972; Harris & Fleer, 1974). Children of 8 years produce RT functions with higher intercepts, but the same slope, as young adults (Hoving et al, 1970; Harris & Fleer, 1974). Introverts are slower than extraverts at scanning for semantic features of category membership (Eysenck & Eysenck, 1979).

Effects of Practice

The effects of extended practice vary with the procedure. If the same fixed set is used over many days, then the RT function becomes flatter and negatively accelerated, particularly when the probe items are consistently associated with one or other response (Ross, 1970; Kristofferson, 1972a). There is some evidence that subjects develop a content-addressable search strategy (Graboi, 1971), and that processing becomes automatic rather than controlled (Shiffrin & Schneider, 1977; Schneider & Shiffrin, 1977). If the memory sets are changed from trial to trial or from session to session, and stimuli are not consistently associated with particular responses, then extended practice affects the intercept but not the slope (Kristofferson, 1972b).

Reliability

The reliability of the Sternberg task has been studied for its possible inclusion in the Performance Evaluation Tests for Environmental Research (PETER) Battery. Twenty-one male subjects performed a 15-minute test session on each of 15 days. Each session comprised five trials requiring an affirmative response and five requiring a negative response at each memory set size from one to four digits presented at the rate of one digit/second. The intercept scores did not change appreciably during the experiment; slopes decreased with practice until the third day, and RT for each of the positive set sizes stabilized after the fourth session. Inter-session reliabilities for both slope and intercept were low, probably because of the small number of trials at each memory set size, but the reliabilities of the RTs from which the slopes were calculated were generally greater than 0.70 (Carter, Kennedy, Bittner, & Krause, 1980; Carter & Krause, 1983).

Split-half reliabilities of the Sternberg task have also been assessed as part of the Taskomat battery (Boer, 1988). The task was administered in two blocks, each of four minutes and comprising approximately 160 trials. In the first block, the memory set was 'R', and in the second it was 'KLMN'. The test stimuli were 2x2 matrices containing either one, two or four letters, and the number of memory comparisons was the product of the memory set and the number of letters in the stimulus array, ie 1, 2 or 4 for "R" blocks and 4, 8 or 16 for "KLMN" blocks. The reliability coefficients were as follows:

	Slope	Intercept
'R' Block	0.32	0.74
'KLMN' Block	0.62	0.65
'R'/'KLMN' Blocks combined	0.76	0.87

A fixed set procedure with two-letter memory sets, similar to the STRES Battery version, was used by Schlegel and Gilliland (in press), who reported reliability of 0.75. Their 123 subjects had previously practised the task for five days, one block of trials per day. The reliability was based on data collected after the practice trials, and the sessions were separated by one day.

Validity

As in the RT task, the question of validity is concerned primarily with the adequacy of the additive-factor framework. Sternberg's finding that RT increases linearly with memory set size, indicating serial search, has been confirmed in several laboratories with different subject samples and levels of practice. However, studies of duplication of items in the memory set, their serial positions, and the relative frequency with which they are tested, have led to disagreement over the type of serial search. Most investigators support the serial exhaustive hypothesis, but several favour the serial self-terminating interpretation, and some prefer a combination of the two.

Support for the discrete-stage information-processing model was provided by Sternberg's finding that degrading the stimuli does not affect the memory search process. With a few exceptions (eg Klatzky et al, 1971), most studies have supported the same model. However, Welford (1980) has argued that the model fails to explain serial order effects.

Sensitivity

The Sternberg task has been used mostly in environmental research to identify the loci of effects of drugs and workload.

Drugs

1. Industrial Chemicals. Smith and Langolf (1981) reported that four levels of exposure to mercury affected the slope but not the intercept of the RT function. Maizlish, Langolf, Whitehead, Fine, Albers, Goldberg, and Smith (1985) reported that long term exposure to mixtures of organic solvents had no effect.

2. Social Drugs. Oborne and Rogers (1983) found that various combinations of alcohol and caffeine affected the intercept but not the slope. Tharp, Rundell, Lester, and Williams (1974) found that alcohol impaired response selection. Roth, Tinklenberg, and Kopell (1977) studied ethanol and marihuana, and reported that the amplitude of the P300 component of the evoked cortical potential showed a drug effect and a set size effect. Both drugs differed significantly from placebo but not from each other, and marihuana increased the overall RT by about 75 milliseconds.

3. Benzodiazepines. Subhan (1984) reported that flunitrazepam and triazolam impaired stimulus encoding and serial comparison stages, whereas lormetazepam had little or no effect. Rizzuto reported that a 5 mg dose of diazepam did not affect performance on this task, whereas a 10 mg dose resulted in significant RT increases but no changes in error scores (Rizzuto, Wilson, Yates, & Palmer, 1985; Rizzuto, 1987)

4. Hypnotics. Rundell, Williams, and Lester (1978) and Williams, Rundell, and Smith (1981) found that secobarbital affected stimulus encoding, but Mohs, Tinklenberg, Roth, and Kopell (1980) reported that it had no effect.

5. Antidepressants. McNair, Kahn, Frankenthaler, and Faldetta (1984) reported that amitriptyline increased performance speed generally by about 7%, but amoxapine had no effect.

6. Stimulants. Naylor, Halliday, and Callaway (1985) reported that methylphenidate speeded response selection but not stimulus evaluation. Mohs et al (1980) reported that methamphetamine had no effect.

7. Anticholinesterases. Wetherell (1986) varied memory set size and stimulus quality and found that physostigmine (previously reported to improve memory) improved stimulus recognition, but not the memory search process.

8. Hormones. Ward, Sandman, George, and Shulman (1979) reported that melanocyte stimulating hormone and adrenocorticotrophic hormone improved stimulus encoding but did not affect memory search rate in men or women.

Workload

1. Dual Tasks. Briggs et al (1972) reported that concurrent performance of a tracking task affected the intercept of the reaction time function but not the slope. Crosby and Parkinson (1979) reported that performance of a ground-controlled approach by pilots affected the intercept but not the slope. Wetherell (1981) reported that car driving appeared to affect the intercept but not the slope for 'yes' responses, and both intercept and slope for 'no' responses. He suggested that subjects were less certain about a 'no' than a 'yes' decision and performed more searches to accumulate confidence before responding.

2. Evoked Cortical Potentials (P300). Gomer, Spicuzza, and O'Donnell (1976) reported that the P300 was enhanced for 'yes' responses, and that the difference in P300 between 'yes' and 'no' responses increased with memory set size. Brookhuis, Mulder, Mulder, Gloerich, van Dellen, van der Meere, and Ellerman (1981) reported that their RT data indicated a self-terminating search process whereas the P300 data indicated an exhaustive search. Adam and Collins (1978) reported that P300 latencies increased with memory set size up to 7 digits, but there were large individual differences and no correlation with set sizes of 9 and 11 digits. Ford, Roth, Mohs, Hopkins, and Kopell (1979) reported that RT was slower in older than in younger subjects, but that there was no difference in P300 latency or amplitude. However, Pfefferbaum, Ford, Roth, and Kopell (1980) reported that the P300 amplitude increased with memory set size and that younger subjects showed larger P300 amplitudes than did older subjects. Rizzuto et al (1985) and Rizzuto (1987) reported that 5 mg of diazepam had no effect on the P300 while a 10 mg dose significantly increased P300 latency and reduced its amplitude.

Simulated Deep-Sea Dives

Lorenz and Lorenz (1988) found that both speed and accuracy of memory search were impaired during simulated dives to maxima of 560 metres of sea water using heliox and 360 metres of sea water using trimix (5% nitrogen).

Technical Specification

Figure 16 illustrates the structure of the task. The Fixed Set procedure is used, and the test is administered in two three-minute blocks, each devoted to one memory set size. This arrangement conforms to the requirements of the tracking task when this and Memory Search are co-administered as a dual task. Two three-minute blocks must be administered to determine the slope and intercept of the RT vs memory set size function. Block 1 uses a memory set size of two items, and Block 2 a set size of four items. Each block is administered separately, and consists of presentation of the memory set followed by a series of probes.

Memory Sets and Probe Items

The memory set letters are randomly selected, without replacement, from all 26 letters of the alphabet. No obviously visually or acoustically confusing letters (eg M and N) are used in the same memory set.

Positive probe letters are equally likely to match any of the memory set letters. Negative probe letters are randomly selected from the letters not used in the memory sets, with the constraint that no negative probe has gross visual or acoustic similarity to any memory set item. The total number of probes presented varies with the subject's RTs, but the order of presentation of positive and negative probes is randomized so that equal numbers are presented on average.

Visual and acoustic confusion depends upon factors such as type-font, language, dialect, and accent. Thus, the composition of memory and probe sets cannot be standardized across cultures. However, the available evidence suggests that, if the test user ensures that confusability is minimized for the subject pool to which the test is administered, the specific choice of items will have negligible effect on test performance.

The elements of the memory set, and the sequence of probe items, should be selected randomly each time a trial block is administered.

Presentation

The memory set letters are presented simultaneously, in a horizontal line across the centre of the monitor screen, with one character space between each letter. Probe letters are presented in the centre of the display area.

Each trial block begins with the presentation of a memory set. The subject views the set for as long as desired, and then removes it by pressing either of the two response keys. The first probe appears one second later, and constitutes the beginning of the three-minute test period. The structure of each trial is as follows: 1) the probe is presented on the screen, 2) as soon as the subject responds, or a deadline of five seconds has elapsed, the probe is erased, 3) the screen remains blank for one second. RTs are measured from the onset of each probe to the first depression of a response key. Thus, if the subject initially makes an incorrect response and immediately attempts to correct it by pressing the other key, RT is calculated to the first response and an error is recorded. After three minutes, the message 'end of block' appears.

Data Specification

A separate data record, listing the memory set and the probes presented, is stored for each three-minute block. With the memory set is recorded the subject's viewing time measured in milliseconds from the presentation of the memory set to depression of either response key. With each probe letter is recorded the subject's RT to that probe, coded as positive for a correct response, negative for an incorrect response, and 0 for a response failure.

Summary statistics are calculated separately for each three-minute block, and comprise a) memory set size; b) memory set inspection time; c) mean of all correct RTs; d) SD of all correct RTs; e) mean of correct RTs to positive probes; f) SD of correct RTs to positive probes; g) mean of correct RTs to negative probes; h) SD of correct RTs to negative probes; i) number of positive trials; j) number of negative trials; k) percent errors on positive trials; l) percent errors on negative trials; m) percent response failures on positive trials; and n) percent response failures on negative trials. In the calculation of error rates (k-l), response failures are excluded.

The following summary statistics are calculated, using linear regression, from the data obtained for each pair of three-minute blocks: a) slope of RT function for positive probes; b) intercept of RT function for positive probes; c) slope of RT function for negative probes; and d) intercept of RT function for negative probes.

Training Requirements

Subjects are given an opportunity to read the instructions, and any questions are answered. They then enter the practice phase, comprising 10 blocks (standard schedule) or two blocks (abridged schedule).

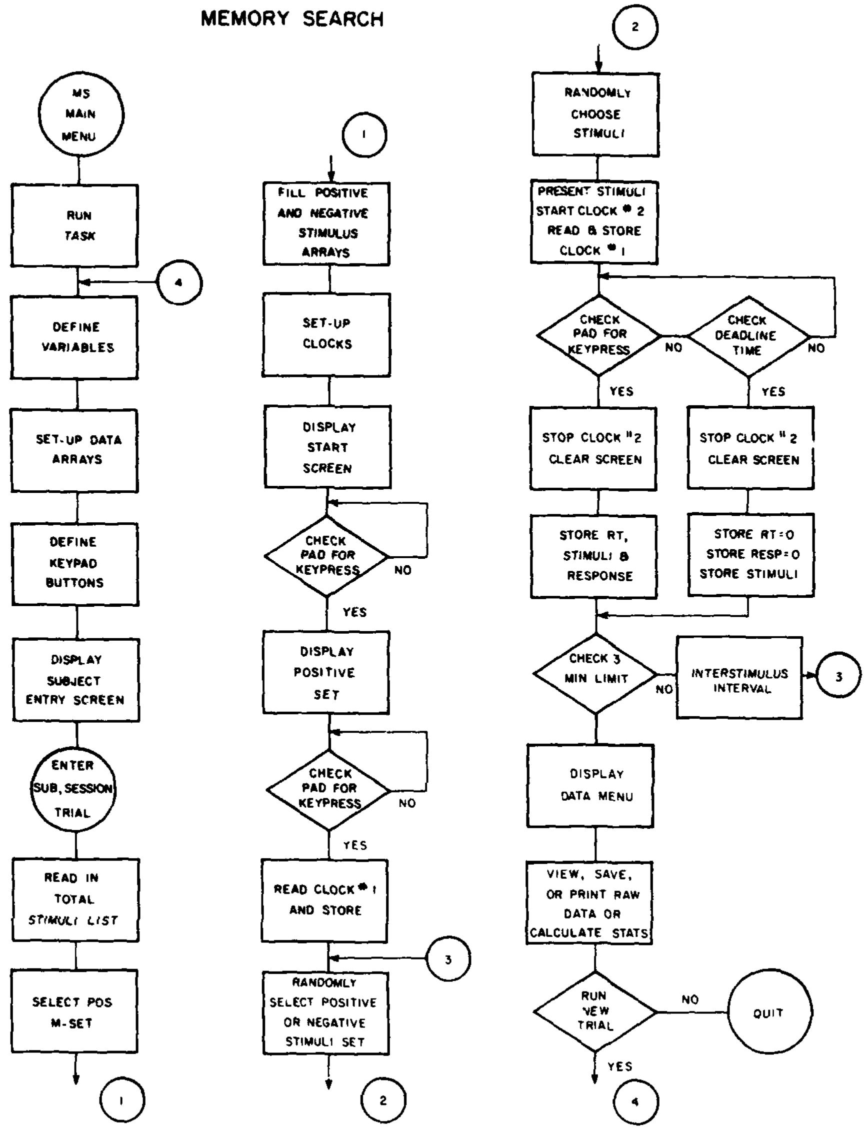


Figure 16. The structure of the Memory Search task.

Instructions to Subjects

This is a test of your ability to search your memory for particular letters. You will be shown a set of letters to memorize, called the "memory set". It will contain either two or four letters, and you will be allowed to look at it for as long as you wish. When you have memorized this set, you should press one of the response keys and you will then be shown a series of single test letters, one at a time. You have to decide whether each test letter is one of the letters in the memory set. If so, press the 'yes' key; if not, press the 'no' key. Please try to respond as fast as you can without making any mistakes. If you do not respond within a certain time, the next letter will appear. Each period devoted to a particular memory set will last for three minutes. Each memory set will be different, so be sure to memorize it before you press the key to start the series of test letters.

SPATIAL PROCESSING TASK

Purpose

This task is designed to examine the subject's ability to rotate histograms mentally prior to making a same/different judgment. It taps visual short term memory, since the standard and test stimuli are presented successively rather than simultaneously.

General Description

On each trial, a pair of four-bar histograms is presented sequentially on the monitor screen. The subject must determine whether the second, 'test', histogram is identical to the first, 'standard', histogram, regardless of an orientational difference of 90 degrees or 270 degrees, and respond 'same' or 'different' by pressing the appropriate response key.

Background

This task is adapted from the spatial processing task used in the CTS (Shingledecker, 1984), which is, in turn, derived from an earlier task devised by Fitts, Weinstein, Rappaport, Anderson, and Leonard (1956) and later used by Chiles et al (1968).

In the STRES version of the task, a standard stimulus oriented at zero degrees is presented, followed, after an interval, by a single test stimulus rotated through 90 or 270 degrees. The test stimulus may be the same as, or different from, the standard stimulus. The standard must be maintained in memory, and the test stimulus mentally rotated prior to the same/different judgment (see Cooper & Shepard, 1978). Thus, storage, transformation, and comparison of visuo-spatial material are all required.

This general paradigm is known as the Fitts Histogram procedure. Fitts and his colleagues presented a single histogram to their subjects as a standard, followed by six rows of eight simultaneously presented test stimuli. The subject's task was to select the test stimulus from each row that was identical to the standard. Some of the stimuli were created in the same fashion as those in the STRES task, using six bars with lengths from one to six units. Others were created as the figure and its mirror image, joined at the midline. And finally, a third group comprised two repetitions of the pattern in the same orientation. In general, Fitts found that RT was fastest for random stimuli, and slowest for constrained stimuli in which the bars were chosen without replacement from the population of possible heights. Moreover, symmetrical stimuli were identified most quickly.

The stimuli used in the present task correspond to Fitts et al's definition of constrained figures, since each bar in the histogram is selected without replacement from a population of all possible bar heights with the result that no two bars have the same height. Fitts and his coworkers found that detection times for such figures were slower than those for random figures in which bars of the same height were permissible.

The Spatial Processing task can be classified as one of spatial transformation, as defined in Lohman's (1979) survey and re-analysis of the correlational literature on spatial ability. More specifically, it requires the visualization (Vz) ability involved in mental reorientation of complex figures. Other, more fundamental, elements of Lohman's classification addressed by this task include perceptual speed (Ps) in the stimulus-comparison component of the task, and perhaps closure speed (Cs), which refers to the speed of matching incomplete or distorted stimuli with representations stored in memory.

Reliability

Kennedy, Dunlap, Jones, Lane, and Wilkes (1985), who used the Fitts Histograms as a paper-and-pencil 'marker' test during the development of a microcomputer-based repeated measures test battery, found a test-retest reliability of 0.90 for data collected on two separate days with one intervening nontest day. Since performance on paper and pencil tests tended to stabilize more slowly than the same test in computer based form, this estimate of reliability is probably conservative.

Chiles et al's (1968) spatial processing task produced a split-half reliability of 0.75. A reliability coefficient of 0.67 was reported on the STRES difficulty level of the Spatial Processing task by Schlegel and Gilliland (in press). The reliability was calculated on data collected for 123 subjects on two separate days following five days of practice, one block per day. The test days were themselves separated by one day.

Validity

Kennedy et al reported that scores on the Fitts Histogram test correlated 0.71 with those on Klein and Armitage's (1979) pattern comparison task. Moreover, Histogram scores loaded onto the same factor as other tests with spatial components, including the Manikin test (related to Lohman's Spatial Orientation factor), and Code Substitution and the Klein and Armitage task (both related to Lohman's Spatial Relations factor). The Histograms also loaded onto a motor control factor, perhaps because the test was administered in paper-and-pencil format. One of the remaining factors had loadings on the computer-based tasks but not their paper-and-pencil counterparts. This finding suggests that fundamentally different stategies may be applied to different versions of the same test, and emphasizes the importance of standardization.

Since Kennedy et al's factor analysis was performed on data for 11 tests obtained from only 20 subjects, the results must be considered tentative. Nevertheless, they are consistent with the notion that histogram comparison taps spatial processing resources.

Sensitivity

Tentative evidence of the sensitivity of this task can be inferred from findings using tests with which it is correlated. The Manikin test, for example, is sensitive to the effects of diving to extreme depth (Lewis & Baddeley, 1981; Logie & Baddeley, 1983); the Klein and Armitage test is sensitive to cyclical variations in arousal (Klein & Armitage, 1979); and a test resembling the Fitts Histograms has been found to reflect the effects of long-term isolation (Chiles et al, 1968; Chiles et al, 1969).

Rizzuto (1987) has reported that a 10 mg dose of diazepam significantly increased RTs but had no effect on percent correct. He further reported that evoked potentials recorded from the task stimuli showed increased P300 latencies and reduced P300 amplitudes. Since the error scores were unchanged it was concluded that the task was performed correctly with the diazepam but that the amount of time required by the stages of processing leading to the responses was increased by the 10 mg dose.

Technical Specification

The structure of the task is depicted in Figure 17. Each histogram comprises four bars one to six units in height, each unit being 8.5 millimetres high and five millimetres wide; adjacent bars are separated by a gap of five millimetres, with a line extending along the base of the figure. The height of each bar in a given histogram is determined randomly, with the constraint that no two bars are identical. A number is presented with each histogram to indicate whether it is a standard stimulus (1) or a test stimulus (2). Standard stimuli are presented in the zero degree orientation with the baseline under the histograms positioned in the middle of the horizontal axis of the screen and 35 millimetres below its centre. The histogram bars extend above the horizontal baseline and the number 1, indicating a standard stimulus, is positioned with its base 50 millimetres below the centre of the screen. For the test stimuli, the histogram extends left (90 degree orientation) or right (270 degree orientation) of screen centre, the centre of the baseline being coincident with the centre of the screen. The number 2, indicating a test stimulus, appears with its base 45 millimetres below the centre of the screen (Figure 18).

The task is performed in three-minute trial blocks. On each trial, the subject must decide whether the test stimulus is identical to the standard stimulus, regardless of difference in orientation, and respond by pressing the 'same' or 'different' key.

The structure of each experimental trial is as follows: 1) the standard stimulus is presented for three seconds; 2) the screen is blanked for one second; 3) the test stimulus is presented; 4) as soon as the subject presses one of the response keys, or a deadline of 15 seconds has elapsed, the test stimulus is erased and a one-second inter-trial interval begins.

Practice trials differ from the experimental trials as follows: 1) as soon as a response is made, the test stimulus is erased, and feedback concerning accuracy and RT is presented on two lines in the middle of the screen; 2) this feedback remains on the screen until the subject presses either response key to initiate the inter-trial interval.

During each three-minute trial block, test stimuli are equally likely to be rotated through 90 or 270 degrees relative to the standard; at each of these orientations, the test stimulus is equally likely to be 'same' or 'different' with respect to the standard. On 'different' trials, the standard and test stimuli must differ by at least one unit on at least one of the component bars.

Data Specification

For every trial within a three-minute trial block, RT (coded as positive for a correct response, negative for an incorrect response, and 0 for a response failure) is recorded.

The following summary statistics are determined for each three-minute block: a) mean of all correct RTs; b) SD of all correct RTs; c) mean of correct RTs for response 'same'; d) SD of correct RTs for response 'same'; e) mean of correct RTs for response 'different'; f) SD of correct RTs for response 'different'; g) number of 'same' trials; h) number of 'different' trials; i) percent errors on 'same' trials; j) percent errors on 'different' trials; k) percent response failures on 'same' trials; and l) percent response failures on 'different' trials. In the calculation of error rates (i-j), response failures are excluded.

Training Requirements

Subjects are given the opportunity to read the instructions, and then complete 10 practice blocks (standard schedule) or 2 practice blocks (abridged schedule). If the task is administered to the same subject in more than one session, practice should be omitted after the first session.

Instructions to Subjects

Practice blocks

In this task, a pair of bar graphs, or histograms, is presented one at a time on each trial. Your task is to memorize the shape of the first of the two histograms, and then decide whether the shape of the second histogram is the same or different. The first histogram is labelled with a "1" and the second with a "2" so that you will not confuse them. Always memorize the shape of the first histogram and press the 'same' or 'different' key, as appropriate, when the second histogram is displayed.

Every histogram will contain four bars. The first of each pair will be presented in an upright position, but the second will be rotated on its left or right side. You should ignore this difference in orientation when deciding whether or not the histograms are identical in shape.

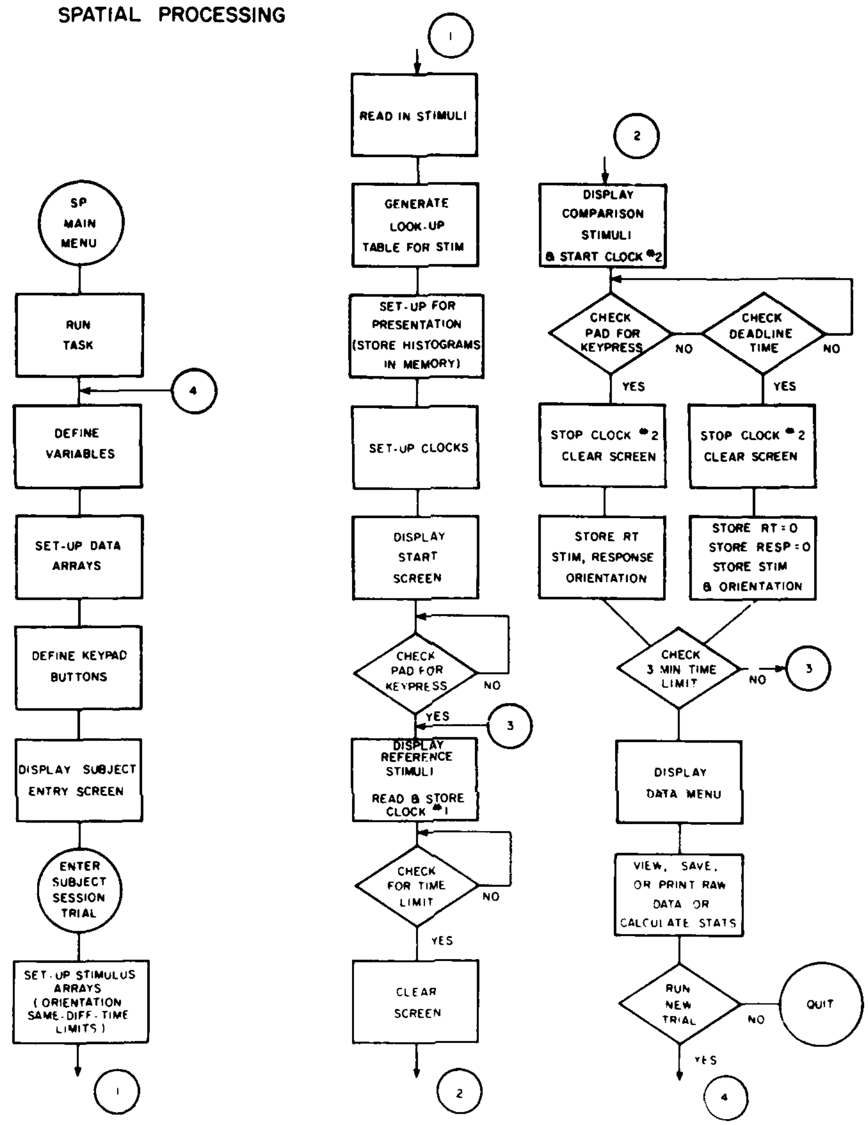


Figure 17. The structure of the Spatial Processing task.

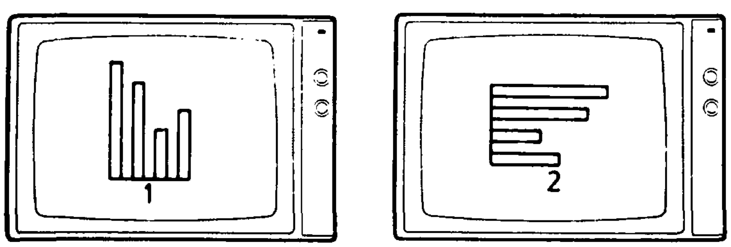


Figure 18. Example of standard and test stimuli in the Spatial Processing Task.

Please start the task whenever you are ready by pressing either of the response keys. Memorize the shape of the first histogram and respond either 'same' or 'different' to the second. As soon as you respond, you will be informed of your reaction time and accuracy. When you are ready to proceed to the next trial, press either of the response keys; the display will be erased and the next problem will appear shortly afterwards. Try to respond as quickly and accurately as possible. In other words, respond as quickly as you can, but if you start making errors because you are rushing your decision, slow down. After three minutes, the message 'end of block' will appear.

Experimental blocks

In this task, a pair of bar graphs, or histograms, is presented one at a time on each trial. Your task is to memorize the shape of the first of the two histograms, and then decide whether the shape of the second histogram is the same or different. The first histogram is labelled with a "1" and the second with a "2" so that you will not confuse them. Always memorize the shape of the first histogram and press the 'same' or 'different' key, as appropriate, when the second histogram is displayed.

Every histogram will contains four bars. The first of each pair will be presented in an upright position, but the second will be rotated on its left or right side. You should ignore this difference when deciding whether or not the histograms are identical in shape.

Please start the task whenever you are ready by pressing either of the response keys. Memorize the shape of the first histogram and respond either 'same' or 'different' to the second. As soon as you respond, the display will be erased and the next problem will appear shortly afterwards. Try to respond as quickly and accurately as possible. In other words, respond as quickly as you can, but if you start making errors because you are rushing your decision, slow down. After three minutes, the message 'end of block' will appear.

UNSTABLE TRACKING TASK

Purpose

This task tests information processing resources used in the execution of continuous manual control responses.

General Description

A fixed target is presented in the centre of the monitor screen. The subject manipulates a joystick in an attempt to maintain the position of a horizontally-moving cursor on the target. The system is inherently unstable: operator input introduces error that is magnified such that it becomes increasingly necessary to respond to the velocity as well as the position of the cursor.

Background

This task was developed by Jex, McDonnell, and Phatak (1966). It was inspired by analytical treatment of aircraft handling qualities, such as Ashkenas and McRuer's (1959) work on just-controllable aircraft short-period static instability and its strong relationship with operator (pilot) effective time delay. Ashkenas and McRuer showed that increased rate of system error associated with control tasks produces corresponding increases in the operator's internal delay in processing and responding to the disturbance. Subsequently, it was reported that control loss occurred at the same static instability level for three test pilots (Jex & Cromwell, 1961). These findings resulted in a more extensive investigation of the dynamics of manual control behaviour, and provided the impetus for the development of a reliable, internally valid control task for applied research. Jex et al (1966) set out to develop such a task and to validate experimentally the assumptions underlying a model of human control behaviour.

Since tracking involves input, translation, and output mechanisms, it has been modelled using techniques borrowed from Fourier analysis and linear feedback control theory. Tracking performance can be described reasonably well by the linear differential equations, or 'transfer functions', incorporated into a quasilinear class of model of the human operator. In quasilinear models, man's response to tracking input signals, although nonlinear, is approximated by a linear transfer function called the 'describing function' and a separate nonlinear component called the 'remnant'. The strength of such models is that their parameters, such as time delay and gain, seem to correspond to specific characteristics of human control behaviour in man-machine systems.

McRuer and Jex's (1967) 'crossover model' is an example of the quasilinear approach. A describing function with the two parameters of effective time delay and gain is used to model the proportion of the subject's response that is linearly correlated with the input signal (Figure 19). This describing function takes the form

$$o(t) = Kse(t - te)$$

where $o(t)$ represents the subject's output at time (t)
Ks represents the subject's gain
te represents the subject's effective time delay in processing the tracking signal
$e(t - te)$ represents the input to the subject, or system error, at time $(t - te)$.

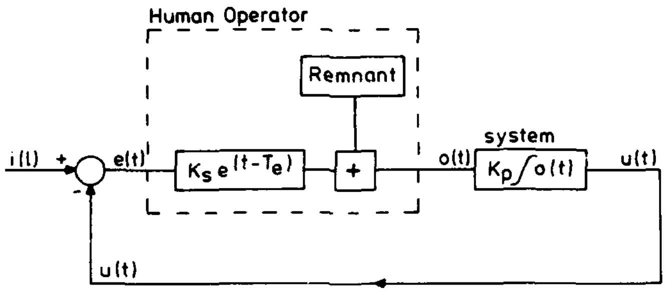


Figure 19. Block diagram of Quasilinear Crossover Model.

The effective time delay (te) has been found to be somewhat analogous to discrete reaction time (Wickens, 1976): it is simply the interval between the introduction of system error and the emission of an appropriate response by the subject. The gain parameter, Ks, is a measure of how large a corrective movement the subject will make in response to a given system error. Subjects who exhibit high Ks values tend to make relatively large amplitude control movements, leading to more oscillatory tracking behaviour under some circumstances. Practised subjects are able to adjust their gain to specified levels. Gain can be considered analogous to response bias, controlled by high-level cognitive processes (Wickens, 1976).

The key characteristic of the Unstable Tracking task is the positive feedback loop responsible for the inherent instability of the system. Once the system detects a control error, it generates an error velocity whose value is determined by operator gain. Unlike systems based on negative feedback, in which this velocity is subtracted from the control error, positive feedback adds the velocity to the error, increasing the rate of error movement away from the target. Thus, the subject's gain adds to the rate of system error, and precise corrective movements are critically important. The dynamics of the Unstable Tracking task are analogous to those of a balanced stick (Wickens, 1984). If an error from the vertical is introduced, the stick will begin to fall, and the rate of falling (increase in error) will increase as it falls.

Although the human operator is better designed to deal with the properties of a negative feedback system, positive feedback loops are characteristic of many complex dynamic vehicles, and demand of the operator constant attention. It is therefore important to understand the inter-relationships between the elements of the describing functions associated with unstable tracking.

The precise parameters of the Unstable Tracking task were determined empirically during Shingledecker's (1984) test development phase. On the basis of two measures of tracking performance (average absolute tracking error and number of control losses), and subjective difficulty ratings, three reliably different demand levels were produced by lambda (instability) values of 1.0 (low demand), 2.0 (moderate demand), and 3.0 (high demand).

This task is assumed to tap primarily motor output resources, placing minimal demands upon resources associated with input and central processing. Evidence for this assumption was provided by Shingledecker, Acton, and Crabtree (1983), who required subjects to perform unstable tracking, at each of three demand levels, concurrently with an interval production task. Interval production variability increased systematically as a function of tracking task demand, but was unaffected by tasks tapping input or central processing stages. It was therefore concluded that unstable tracking and interval production place demands primarily upon resources devoted to motor responses.

Reliability

The reliability and stability of critical tracking scores (degree of instability when control is lost) vary with practice. Damos, Bittner, Kennedy, and Harbeson (1981) examined the critical tracking performance of 12 subjects during 15 sessions. Performance was found to stabilize after 10 sessions. The mean correlation between performance on the final five sessions was 0.764.

In Damos, Bittner, Kennedy, Harbeson, and Krause's (1984) study, in which subjects performed the task on 14 days, tracking performance based on critical instability scores became relatively stable after 105 brief practice trials. Although slow linear improvement in scores was apparent from day 8 until the end of the testing period, it was concluded that the task is sufficiently reliable for use in dual-task, environmental stress, or drug studies, provided that proper attention is given to the effects of practice.

A reliability of O.83 has been reported by Schlegel and Gilliland (in press) for the mean absolute error in the STRES version of the Unstable Tracking task. They also reported a reliability coefficient of O.82 for the number of edge violations (control losses) in this task. Their sample comprised 12O male and female subjects who had practised the task for five days, one block of trials per day. The reliability data were collected on two days, with one intervening day, following training.

Validity

Jex et al (1966) concluded that there is "good experimental validation of the theoretical assumptions and implications of the operator's behaviour (with respect to the elements of a describing function) in the first order critical task" (p. 142). These authors used the three-parameter Extended Crossover Model (ECM) of McRuer, Graham, Krendel, and Reisner (1965) to fit the data, and established an operator describing function.

Experimental evidence indicates that the effective time delay (te) approaches an irreducible minimum and flattens out as extreme instability (system error) is reached (see Jex et al. 1966, Figure 4A), and that the gain margin (the gain necessary to prevent the subject from lagging 180 degrees or more behind the system) decreases as instability increases. Actual operator gain closely follows the theoretical gain for the maximum gain margin delineated by the describing function; gain limitations are constrained as critical limits are approached. These findings concerning the effects of instability on operator gain and effective time delay conform closely to the predictions of the ECM, and hence indicate high construct validity.

Sensitivity

Klein and Jex (1975) showed that alcohol produced a decrement in critical tracking performance. Similarly, Dott and McKelvy (1977) reported that mean tracking error and total error increased, whereas degree of instability when control was lost decreased, as a function of blood alcohol level. Moreover, the effects of secobarbitol and carbon monoxide on positive feedback tracking have been found to be similar to those of alcohol (Putz, 1979). It appears that inherent instability may be necessary for tracking tasks to reveal the effects of drugs and toxic substances. Klein and Jex, for example, noted that traditional negative-feedback tracking tasks show little sensitivity to the effects of alcohol.

Lorenz and Lorenz (1988) reported that unstable tracking performance declined substantially during simulated deep-sea dives, but recovered rapidly during subsequent decompression.

Extensive research on the effects of acceleration (G-stress) on tracking has been conducted at the Armstrong Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio. Although the magnitude of these effects is influenced by factors such as task dynamics, direction of acceleration, subject position, and use of G-force protective suits (see reviews by Grether, 1971; Little, Hartman, & Leverett, 1968; Van Patten, 1984), it is well-established that unstable tracking is sensitive to variation in G.

Jex, Peters, DiMarco, and Allen (1974) hypothesized that physiological deconditioning from orbital living (in the form of 10 days of enforced bedrest) might degrade the pilot's ability to control his aircraft manually during shuttle reentry. Forty-two subjects, each provided with a G-suit, were subjected to acceleration before and after bedrest. Although bedrest had no overall effect on mean critical scores, it interacted with centrifugation. Before bedrest, critical tracking following a centrifuge run was non-significantly better than that prior to the run; after bedrest, however, 62 percent of the post-run scores were worse than pre-run scores. Thus, it appears that the enforced bedrest interfered with G-protected subjects' ability to overcome the effects of acceleration.

Adler, Strasser, and Muller-Limmroth (1976) showed that tracking performance on a task resembling that devised by Jex was superior under distributed relative to massed practice, was degraded when the practice regime was changed, and was improved by monetary incentive.

A 10 mg oral dose of diazepam has been shown to increase tracking error and the number of edge violations. These effects were reported for two levels of difficulty of the critical tracking task. Evoked potentials were recorded to offset blinks of the tracked cursor, and showed latency increases and amplitude decrements in the P300 (Rizzuto, 1987).

Schlegel and Gilliland (in press) showed that tracking performance was significantly impaired by one night's sleep loss. Their subjects performed three levels of the unstable tracking task, including the STRES Battery level. Absolute mean tracking error, but not the number of edge violations, was adversely effected by this level of sleep loss.

Technical Specification

The structure of the task is illustrated in Figure 20. Although detailed consideration of the mathematical characteristics of the task is inappropriate here, it may be noted that the unstable plant dynamics are a first-order divergent element of the form:

$$P(s) = \frac{\text{lambda}.\exp(-ts)}{s - \text{lambda}}$$

where: P(s) = ratio of system output to input
s = Laplace operator (indicates system response is a function of frequency)
lambda = level of instability = 1/T(seconds), where T (seconds) is divergent time constant
exp(-ts) = Additional phase lag produced by time delay, t

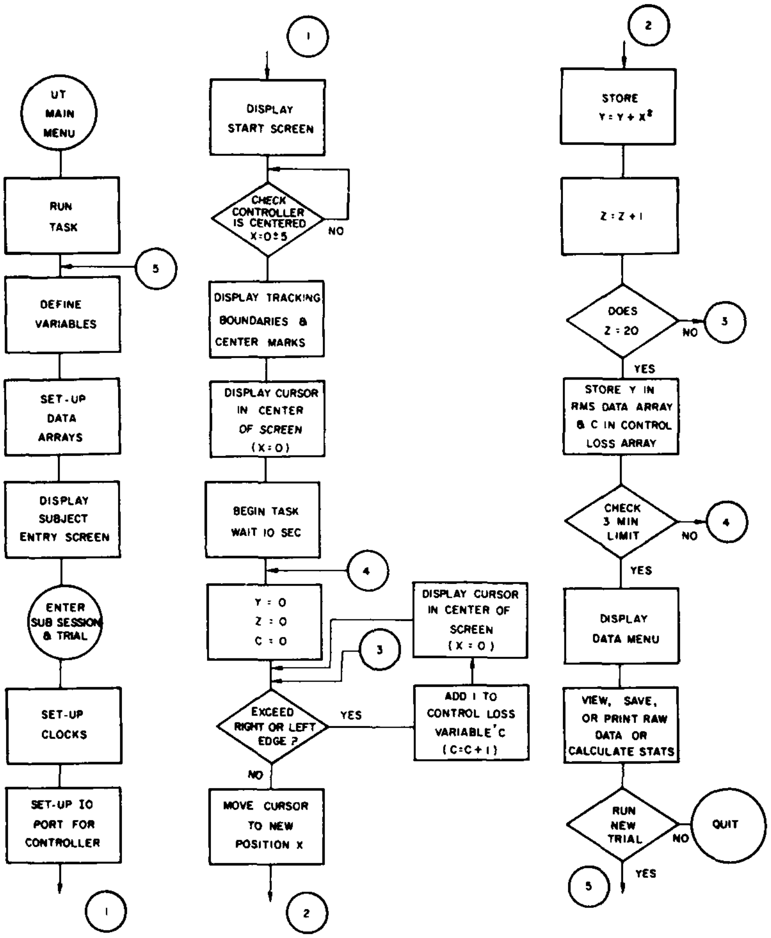
An analogue-to-digital value of zero is obtained when the joystick is centralized, and positive and negative values obtained when the joystick deviates right and left, respectively, from the central position. The task begins as soon as the subject has manipulated the joystick to select a value of zero, using visual feedback of analogue-to-digital converter values displayed in the centre of the screen. The subject is then given 10 seconds to gain control of the cursor before data collection commences.

The position of the cursor on the screen is determined by the following relationship:

New Position = (2 * Rate + Lambda) * Old Position / (2 * Rate — Lambda) + Lambda * Gain * (Stick Input + Last Stick Input) / (2 * Rate — Lambda)

where, for the STRES battery, Rate = 50 Hz
Lambda = 2
Gain = 4

UNSTABLE TRACKING



X = CURSOR POSITION
Y = CUMULATIVE SQUARED ERROR
Z = SAMPLE NUMBER (ERROR IS SAMPLED AT 20 Hz)
C = CONTROL LOSSES

Figure 20. The structure of the Unstable Tracking task.

Figure 21 shows the screen display for the Unstable Tracking task. The tracking cursor moves horizontally, its central position coinciding with the middle of the display screen. The cursor is 15 millimetres high and 2 millimetres wide, with a horizontal bar (2 millimetres high and 5 millimetres wide) at its centre. Screen centre markers, each 2 millimetres wide and 8 millimetres high, are positioned above and below the cursor in the middle of the screen; when the cursor is centred, it forms a continuous line with these markers. Edge markers appear 70 millimetres left and right of screen centre, providing a 140-millimetre tracking range that should include at least 147 screen pixels with a pixel width of no more than 0.95 millimetre. The edge markers are two millimetres wide by 15 millimetres high. Some computers may require EGA graphics capability for sufficient screen resolution.

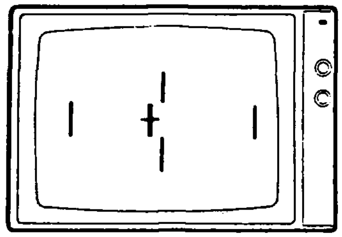


Figure 21. The screen display for the Unstable Tracking task.

Subjects are instructed to attempt to maintain the position of the cursor between the two centre markers throughout the tracking period, avoiding control losses in which the cursor reaches the edge of the screen.

The maximum tracking loop time delay of 5O milliseconds must be accurate to within 5%. No external forcing function is applied to the tracking loop; the unstable dynamics of the task are excited exclusively by human tracking error and by noise in the joystick digitization process. There must be some noise in the digitizing process, and parameters may have to be adjusted to provide this noise. If the subject loses control and the cursor travel reaches the edge of the display, a control loss is recorded, the cursor is automatically re-positioned at the screen centre, and the subject continues tracking. After three minutes, the task ends and the message 'end of block' appears.

Data Specification

The following data are stored for each one-second interval of the task: 1) average error, and 2) incidence of control failure.

Summary statistics for the complete three-minute period comprise a) RMS error score, calculated as:

RMS error = square root ((sum(x) ^ 2) — ((sum(x)) ^ 2 / n) / n — 1) where x = the deviations from screen centre summed for each second, and n = 180

b) The number of control failures.

Training Requirements

Studies by Damos et al (1981) and Shingledecker (1984) indicate that a standard training schedule of 10 three-minute blocks should be adopted to achieve stability of performance. The abridged schedule for this task is two blocks.

Note that, if the task is administered to the subject in several sessions, practice should be omitted after the first session.

Instructions to Subjects

In this task, your objective is to keep a cursor centred on a target area in the middle of the monitor screen. You can control the movement of the cursor by moving the joystick. Moving the stick to the right moves the cursor to the right, and moving it to the left moves the cursor to the left. The cursor initially appears on the central target but tends to move horizontally away from this position. Try to keep it centred over the target at all times. If it reaches the boundary line, it will reappear at the target position and begin moving away again. This is called a control loss and should be avoided if possible.

To begin, please move the joystick until the numerical display on the screen reaches zero. After about three minutes, the message 'end of block' will appear.

GRAMMATICAL REASONING TASK

Purpose

This task, derived from that described by Baddeley (1968), addresses the ability to manipulate grammatical information, placing demands primarily upon working memory.

General Description

On each trial, two sentences with an active and positive construction are presented, together with three symbols; the subject must compare the veracity of the description of the order of symbols contained in the sentences.

Background

Several types of grammatical reasoning tasks have been reported in the literature. Below, five of these procedures are considered.

Wason (1961) presented sentences describing a number as odd or even, such as 'seventy-six is an even number' (true affirmative) or 'seventy-six is not an odd number' (true negative). Combinations of the factors of true/false and affirmative/negative were used to generate 24 stimuli. Wason found that negative statements were verified more slowly than positives. He suggested that this difference reflected the additional time required to invert the negative form (eg 'not even') to positive (eg 'odd').

The advantage for positive sentences was confirmed using other techniques. Slobin (1966) presented a sentence followed by a picture (eg a cat chasing a dog); subjects were required to decide whether the sentence correctly described the picture. Clark and Chase (eg Chase & Clark, 1972; Clark & Chase, 1972, 1974) required subjects to compare the * + sentence 'the star is not above the plus' to stimuli such as + (false) or * (true). In both of these paradigms, as in Wason's, it appeared that a time-consuming process of inversion was occurring for negative sentences.

Baddeley's (1968) grammatical reasoning task was inspired by the findings reported by Slobin (1966) and Wason (1961). In this task, a statement describing the order of letters A and B was accompanied by the letter pair AB or BA (eg B is not followed by A — BA); subjects were required to indicate whether or not the statement correctly described the letter pair. Thirty-two different problems were generated by combination of 1) use of the verb 'precede' or 'follow'; 2) active or passive voice; 3) affirmative or negative construction; 4) order of A and B in the statement; and 5) order of A and B in the letter pair. In this task, affirmative sentences were verified more quickly than negative sentences, and active more quickly than passive.

Baddeley and Hitch (1974) and Hitch and Baddeley (1976) showed that a concurrent memory load of six letters slowed verbal reasoning performance but had no effect upon accuracy. Thus, it appeared that the short-term memory store and the system responsible for reasoning were at least partially overlapping. There is little doubt that verbal reasoning places demands upon central resources. Farmer, Berman, and Fletcher's (1986) finding that articulatory suppression (repetition of irrelevant speech sounds) interferes with verbal, but not spatial, reasoning suggests also the involvement of the specialized verbal subsystem known as the 'articulatory loop' (see, for example, Baddeley, 1986).

Shingledecker (1984) substituted the symbols used by Clark and Chase for the letters A and B within Baddeley's task. The STRES version continues the use of symbolic rather than alphabetic stimuli, but departs more dramatically from the original technique by avoiding the use of the passive voice, which is seldom used in German and might therefore be responsible for cultural differences in test performance. In an attempt to redress the reduction in difficulty caused by elimination of passive stimuli, two statements specifying the order of three symbols are presented on each trial of the STRES task.

Reliability

Baddeley (1968) reported reasonably high test-retest reliability for his test, which was administered in paper-and-pencil form. He tested 18 subjects twice on successive days, and found that the average correlation between performance on the two days was 0.80.

Carter, Kennedy, and Bittner (1981) examined the reliability of a grammatical reasoning test similar to Baddeley's but reduced in duration from three minutes to one minute. Thirty-six subjects were tested on 15 consecutive work-days. Using as a performance measure number of correct responses within each one-minute period, Carter et al found that a) performance increased linearly with practice; b) the variances were stable over the 15 days of testing; c) inter-trial correlations tended to remain constant, especially after the fourth day of testing; and d) the average inter-trial correlation after day 4 was 0.82. These results, together with those of Baddeley (1968), indicate not only that the paper and pencil version of the traditional grammatical reasoning task is a reliable instrument, but also that it is robust to procedural variations such as reduction of test duration that often decrease test reliability. The STRES Grammatical Reasoning task differs from that described by Baddeley in several respects, and its reliability remains to be established. However, since there is clearly considerable overlap between the processes tapped by these tasks, the STRES version is likely to exhibit adequate reliability.

Schlegel and Gilliland (in press) tested the reliability of the CTS version of the grammatical reasoning task and reported a reliability coefficient of O.83. They employed 123 subjects who had practised the task for one block of trials per day for five days, and tested the reliability of data collected on two subsequent days, the latter separated by one day. Although this task is not identical to the STRES version, it is very close in construction and provides at least an estimate of the reliability that should be expected for the STRES Battery version.

Validity

Baddeley (1968) reported a correlation of 0.59 between performance on his grammatical reasoning task and scores on the British Army Verbal Intelligence Test (n = 29). Carter et al (1981) obtained a correlation of 0.44 between grammatical reasoning and the Wonderlic Test of Mental Ability (n = 23). Wetherell (1976) reported that verbal reasoning performance was not significantly correlated with performance on Raven's Standard Progressive Matrices (r = 0.22; n = 30), a test tapping spatial ability. These findings support the notion that the grammatical reasoning paradigm taps 'higher mental processes' associated with verbal ability.

It has been suggested (Hunt, 1980) that the general ability factor (g) of classical intelligence theory may correspond to central resources in modern information-processing approaches. Within Baddeley's model of working memory (see Baddeley, 1986), verbal reasoning places demands upon the limited-capacity attentional system known as the 'central executive'. Farmer et al (1986) showed that it also loads the specialized verbal subsystem of working memory (the 'articulatory loop') but not the spatial subsystem (the 'visuo-spatial sketch-pad'). Hence, for example, recall of verbal memory loads is more greatly impaired by verbal reasoning than by spatial reasoning (Wetherell, 1984a).

Sensitivity

The sensitivity of the STRES version of the grammatical reasoning paradigm remains to be established. However, the CTS version of this task was found to be affected by sleep loss of 24 hours and two mg/kg and four mg/kg of caffeine. RTs were significantly longer under both stressors.

The traditional grammatical reasoning task is sensitive to the effects of numerous environmental stressors. Kemp and Wetherell (1977) reported that 10 mg oral diazepam significantly impaired performance on this task from 15 minutes to two hours after dosing; Holland et al (1978) found that two mg intramuscular (im) atropine, and two mg im atropine with five mg im diazepam, significantly impaired performance from 30 minutes to four hours after dosing, but that five mg im diazepam alone had no effect; and Wetherell (1984b) found that intravenous physostigmine impaired verbal reasoning performance, but only when a verbal memory pre-load was imposed. Other stressors to which this task is sensitive include nitrogen narcosis (Baddeley, de Figuerado, Hawkswell Curtis, & Williams, 1968) and anxiety prior to decompression (Ussher & Farmer, 1987), but not simulated deep-sea diving (Lewis & Baddeley, 1981; Lorenz & Lorenz, 1988).

Verbal reasoning is impaired when performed concurrently with practical tasks. For example, Brown, Tickner, and Simmonds (1969) reported a 44% decrement in number of verbal reasoning problems attempted, and a 28% decrement in number of correct answers, when subjects were driving and judging whether a gap was wide enough to drive through.

The Yerkes-Dodson Law (Yerkes & Dodson, 1908) suggests that the arousal level associated with optimal performance is inversely related to task difficulty. It appears that task difficulty corresponds to the extent to which temporary storage in working memory is required (Hockey & Hamilton, 1983): tasks such as continuous serial reaction (Leonard, 1959) can be classified as 'easy', and tasks such as grammatical reasoning as 'difficult'. Grammatical reasoning is therefore likely to be more sensitive to stressors that increase the arousal level than to those that produce under-arousal. Thus, Farmer and Green (1985) reported that loss of a single night's sleep had a profound effect on continuous serial reaction, but no effect on verbal reasoning.

Technical Specification

The structure of the STRES Grammatical Reasoning task is shown in Figure 22. In this task, the subject is required to compare the veracity of two sentences describing the order of the two adjacent pairs within a set of three symbols (Table 5). If the sentences have the same truth value (both true or both false), the response 'same' is required; if they have different truth values, the response 'different' is required.

Table 5 shows the 32 stimuli selected for use in the task. These stimuli represent each combination of 1) 'before' or 'after' in first sentence; 2)'before' or 'after' in second sentence; 3) first sentence true or false; 4) second sentence true or false; and 5) mapping of first and second sentence onto first and second adjacent letter pair.

During each three-minute testing session, the 32 problems are presented in newly permutated order. If the subject completes more than 32 problems, this permutated order of presentation is repeated until the end of the testing period is reached. The message 'end of block' is then presented.

The structure of each experimental trial is as follows: 1) the problem is presented in the middle of the monitor screen, as shown in Figure 23; 2) as soon as the subject presses one of the response keys, or a deadline of 15 seconds has elapsed, the problem is erased and a one-second inter-trial interval begins.

Practice trials differ from the experimental trials as follows: 1) as soon as a response is made, the test stimulus is erased, and feedback concerning accuracy and RT is presented on two lines in the middle of the screen; 2) this feedback remains on the screen until the subject presses either response key to initiate the inter-trial interval.

GRAMMATICAL REASONING

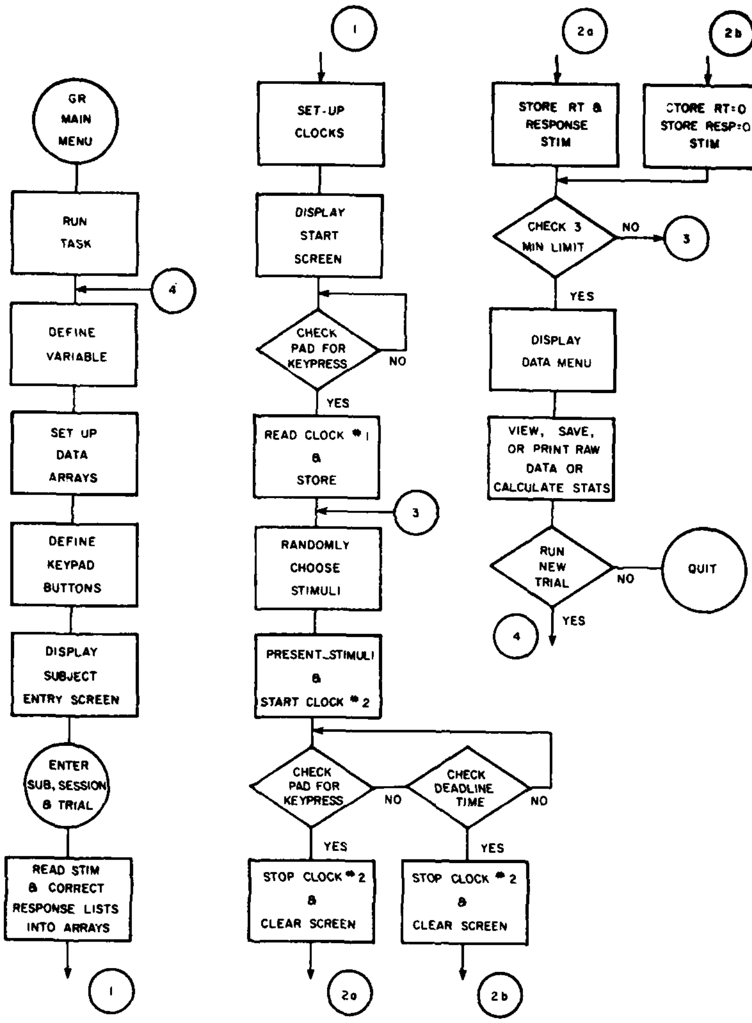


Figure 22. The structure of the Grammatical Reasoning task.

Table 5. Problems used in the Grammatical Reasoning Task. T/F = truth value (true/false) of the sentence relative to the symbol order.

Stim. No.	Sentence 1	T/F	Sentence 2	T/F	Symbol Order	Correct Response
1	* BEFORE &	T	& BEFORE #	T	*&#	same
2	# BEFORE *	T	& BEFORE #	T	&#*	same
3	* BEFORE #	F	& BEFORE *	F	#*&	same
4	* BEFORE &	F	& BEFORE #	F	#&*	same
5	* BEFORE #	T	& AFTER #	T	*#&	same
6	# AFTER *	T	& BEFORE *	T	&*#	same
7	# BEFORE &	F	# AFTER *	F	&#*	same
8	* AFTER &	F	* BEFORE #	F	#*&	same
9	& AFTER *	T	& BEFORE #	T	*&#	same
10	# BEFORE &	T	# AFTER *	T	*#&	same
11	# AFTER &	F	* BEFORE &	F	#&*	same
12	# BEFORE *	F	& AFTER *	F	&*#	same
13	# AFTER *	T	& AFTER #	T	*#&	same
14	# AFTER *	T	* AFTER &	T	&*#	same
15	* AFTER &	F	& AFTER #	F	*&#	same
16	# AFTER *	F	& AFTER #	F	&#*	same
17	# BEFORE *	T	& BEFORE *	F	#*&	diff
18	* BEFORE &	F	# BEFORE &	T	#&*	diff
19	# BEFORE *	F	# BEFORE &	T	*#&	diff
20	* BEFORE #	T	* BEFORE &	F	&*#	diff
21	& BEFORE #	T	# AFTER *	F	&#*	diff
22	& AFTER #	F	* BEFORE &	T	*&#	diff
23	# BEFORE *	F	& AFTER #	T	*#&	diff
24	& AFTER *	T	* BEFORE #	F	#*&	diff
25	& AFTER #	T	* BEFORE &	F	#&*	diff
26	# BEFORE *	F	* AFTER &	T	&*#	diff
27	* AFTER &	F	& BEFORE #	T	*&#	diff
28	* BEFORE &	T	# AFTER *	F	#*&	diff
29	# AFTER &	T	# AFTER *	F	&#*	diff
30	& AFTER *	F	& AFTER #	T	#&*	diff
31	& AFTER *	F	# AFTER *	T	&*#	diff
32	& AFTER #	T	* AFTER #	F	*#&	diff

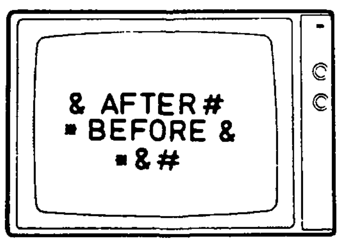


Figure 23. Sample stimulus display for the Grammatical Reasoning task.

Data Specification

For every trial within a three-minute trial block, the following data are recorded: 1) stimulus number (see Table 5); and 2) RT, coded as positive for a correct response, negative for an incorrect response, and 0 for a response failure.

The following summary statistics are determined for each block: a) mean of all correct RTs; b) SD of all correct RTs; c) mean of correct RTs for response 'same'; d) SD of correct RTs for response 'same'; e) mean of correct RTs for response 'different'; f) SD of correct RTs for response 'different'; g) number of 'same' trials; h) number of 'different' trials; i) percent errors on 'same' trials; j) percent errors on 'different' trials; k) percent response failures on 'same' trials; and l) percent response failures on 'different' trials. In the calculation of error rates (i-j), response failures are excluded.

More detailed analysis, such as examination of differences between use of 'before' and 'after', may be undertaken if desired.

Training Requirements

The standard training schedule comprises eight three-minute blocks, and the abridged schedule two blocks. If the task is administered to the subject in several sessions, practice should be omitted after the first session.

Instructions to Subjects

Practice blocks

On each trial, you will be presented with a pair of sentences accompanied by three symbols in a particular order. Each sentence either correctly or incorrectly describes the order of an adjacent pair of symbols within the set of three, and you are required to compare the truth of the sentences. If both sentences are true, or if both are false, press the key marked 'same'; if one sentence is true but the other is false, press the key marked 'different'.

Here is an example:

& BEFORE #
& AFTER *
# & *

The & does not come before the #, so the first sentence is false; and the & does not come after the *, so the second sentence is also false. Since both are false, the correct response is 'same'.

Now examine the following example:

* AFTER &
& BEFORE #
# & *

The * comes after the &, and so the first sentence is true; the & does not come before #, so the second sentence is false. The correct response is therefore 'different'.

You should try to respond as quickly and accurately as you can to each problem. Each time you respond in this practice session, you will be given feedback about your speed and accuracy. When you are ready to begin the next trial, press either response key.

If you find yourself making repeated errors because you are not taking enough time for your decision, slow down. However, do not take more time than is necessary to make the appropriate decision and response.

Please start this practice session by pressing either response key. The session will last for three minutes, after which the message 'end of block' will appear.

Experimental blocks

On each trial, you will be presented with a pair of sentences accompanied by three symbols in a particular order. Each sentence either correctly or incorrectly describes the order of an adjacent pair of symbols within the set of three, and you are required to compare the truth of the sentences. If both sentences are true, or if both are false, press the key marked 'same'; if one sentence is true but the other is false, press the key marked 'different'.

Here is an example:

& BEFORE #
& AFTER *
# & *

The & does not come before the #, so the first sentence is false; and the & does not come after the *, so the second sentence is also false. Since both are false, the correct response is 'same'.

Now examine the following example:

* AFTER &
& BEFORE #
# & *

The * comes after the &, and so the first sentence is true; the & does not come before #, so the second sentence is false. The correct response is therefore 'different'.

You should try to respond as quickly and accurately as you can to each problem. Each time you respond, the problem will be erased and the next problem will be presented after a brief delay.

If you find yourself making repeated errors because you are not taking enough time for your decision, slow down. However, do not take more time than is necessary to make the appropriate decision and response.

Please start this session by pressing either response key. The session will last for three minutes, after which the message 'end of block' will appear.

TRACKING WITH CONCURRENT MEMORY SEARCH

Purpose

This combination of the Unstable Tracking and Memory Search tasks measures the ability to divide attention between two activities.

General Description

During concurrent presentation of these tasks, each proceeds as previously described. Thus, the first three-minute period is devoted to a memory set of two, and the second to a memory set of four. Subjects are instructed to allocate equal priority to the tracking and memory search tasks.

Background

For a task requiring a given type of central processing, some mappings of input and output modalities are more efficient than others (Greenewald, 1979). Wickens, Vidulich, Sandry, and Schiflett (1981) argued, for example, that auditory input and vocal output represent a particularly compatible arrangement for verbal tasks, whereas visual input and manual output are appropriate for spatial tasks.

Vidulich and Wickens (1981) combined tracking with a memory search task. Memory search stimuli were presented either visually or auditorily, and subjects responded vocally or manually. It was found under both single-and dual-task conditions that memory search was performed best with auditory input and vocal output, and most poorly with visual input and manual output. It appeared that there was little central interference between the spatial tracking task and the verbal memory search task: tracking difficulty exerted a negligible effect on memory search performance provided that the tasks were assigned different input and output modalities. When both tasks were presented visually, memory search was more severely disrupted; when both required manual responses, however, degradation occurred primarily in tracking performance. Thus, memory search may impose greater demands on input-related resources, and tracking on response-related resources.

Shingledecker et al (1983) combined a tapping task (Michon, 1966) with other tasks, including tracking and memory search. The Michon task interfered with tracking, but had no effect upon memory search performance. Since the Michon task is assumed primarily to tap resources associated with response timing, this pattern of dual-task interference supports the hypothesis that tracking places a heavy burden on resources associated with response processing.

Task-hemispheric integrity (Wickens & Sandry, 1982; Wickens, Sandry, & Hightower, 1982) must also be considered in the design of concurrent tasks. The dominant cerebral hemisphere is specialized for verbal processing, and the non-dominant hemisphere for spatial processing; each hemisphere controls the actions of the contralateral hand. Task-hemispheric integrity is therefore achieved when a verbal task is performed with the dominant hand, and a spatial task is performed with the non-dominant hand (Wickens, 1981).

Wickens and Sandry (1982) used verbal and spatial versions of the memory search task in combination with a tracking task. For the verbal task, use of the dominant hand produced better time-sharing efficiency than use of the non-dominant hand. There was evidence that the spatial memory search task and the tracking task competed for similar resources, precluding the possibility of presenting both tasks in an integral configuration.

The STRES task combination employs the memory search configuration (visual input with manual output) most likely to produce dual-task interference due to competition for input and output resources. Moreover, task-hemispheric integrity is low: subjects respond to the verbal memory task using the non-preferred hand, and to the spatial tracking task using the preferred hand.

Reliability

The reliability of each of these tests in isolation has already been discussed. There is no direct evidence concerning their reliability in combination. However, the test-retest reliability of tracking with other concurrent tasks (Wickens, Mountford, & Schreiner, 1980) is encouraging, and there is little reason to doubt that the STRES tracking/memory search combination will prove to be adequate in this respect.

Validity

There has been some attempt to identify a general time sharing factor, with inconclusive results (Wickens et al, 1980; Sverko, 1977; Keele & Hawkins, 1982). Although single-dual task performance differences on these tasks decline with practice (Wickens & Sandry, 1982), it is unclear whether this change reflects improvement in time-sharing ability, or simply reduction in the resource requirements of each task. Regardless of the specific mechanism underlying the division of attention between these tests, the evidence cited above indicates that, in their present configuration, they compete for input and output resources.

Sensitivity

The relatively few investigations of tracking with concurrent memory search have been concerned primarily with the development of theoretical models of mental resources. However, evidence for the sensitivity of these tests in isolation has already been presented; moreover, continuous tracking has successfully been combined with tasks involving discrete reactions in several stressor studies, some of which are discussed briefly below.

Putz and his associates (Putz, Anderson, Setzer, & Croxton, 1981; Putz, 1979; Putz, Johnson, & Setzer, 1979) examined the effects of toxic substances on the performance of tracking with concurrent tone detection. Substances such as carbon monoxide and alcohol impaired tracking performance but did not affect tone detection.

Houghton, McBride, and Hannah (1985) used multiple tasks in the study of loss of consciousness induced by G-stress. Two-dimensional compensatory tracking served as the primary task, and two-choice reaction time and mental arithmetic as secondary tasks. The results indicated significant impairment in choice reaction time and mental arithmetic, but no impairment in the primary tracking task.

Farmer and Green (1985) subjected civil aircrew to loss of a single night's sleep. Their battery of tasks included compensatory tracking performed concurrently with detection of peripheral signals. Performance on both of these tasks declined under sleep loss.

Technical Specification

The structure of the dual-task is shown in Figure 24. The tasks proceed as previously specified, with the following exceptions: The cursor is initially centred under software control. As soon as the subject presses a response key to indicate that he has memorized the memory set, the 10-second warm-up period of the Unstable Tracking task begins. The memory set remains on the screen for the first nine seconds of this period. After the 10 seconds have elapsed, the first probe item is presented and the three-minute memory search and tracking period begins.

Memory sets and probe items are presented directly above the centre of the tracking target, with the base of the letters 22 millimetres above screen centre. Figure 25 depicts the stimulus display.

As under single-task conditions, the first three-minute block is devoted to a memory set of two items, and the second to a memory set of four items. Subjects respond to the memory search stimuli using the non-dominant hand and manipulate the joystick using the dominant hand.

Data Specification

For the Unstable Tracking task, the following data are stored for each one-second interval of the task: 1) average error, and 2) incidence of control failure.

Summary statistics for the complete three-minute period comprise a) RMS error score, calculated as:

RMS error = square root ((sum(x) ˆ ß2) — ((sum(x)) ˆ 2 / n) / n — 1) where x = the deviations from screen centre summed for each second, and n = 180

b) the number of control failures.

For the Memory Search task, a separate data record, listing the memory set and the probes presented, is stored for each three-minute block. With the memory set is recorded the subject's viewing time measured in milliseconds from the presentation of the memory set to depression of either response key. With each probe letter is recorded the subject's RT to that probe, coded as positive for a correct response, negative for an incorrect response, and 0 for a response failure.

Summary statistics are calculated separately for each three-minute block, and comprise a) memory set size; b) memory set inspection time; c) mean of all correct RTs; d) SD of all correct RTs; e) mean of correct RTs to positive probes; f) SD of correct RTs to positive probes; g) mean of correct RTs to negative probes; h) SD of correct RTs to negative probes; i) number of positive trials; j) number of negative trials; k) percent errors on positive trials; l) percent errors on negative trials; m) percent response failures on positive trials; and n) percent response failures on negative trials. In the calculation of error rates (k-l), response failures are excluded.

The following summary statistics are calculated, using linear regression, from the data obtained for each pair of three-minute blocks: a) slope of RT function for positive probes; b) intercept of RT function for positive probes; c) slope of RT function for negative probes; and d) intercept of RT function for negative probes.

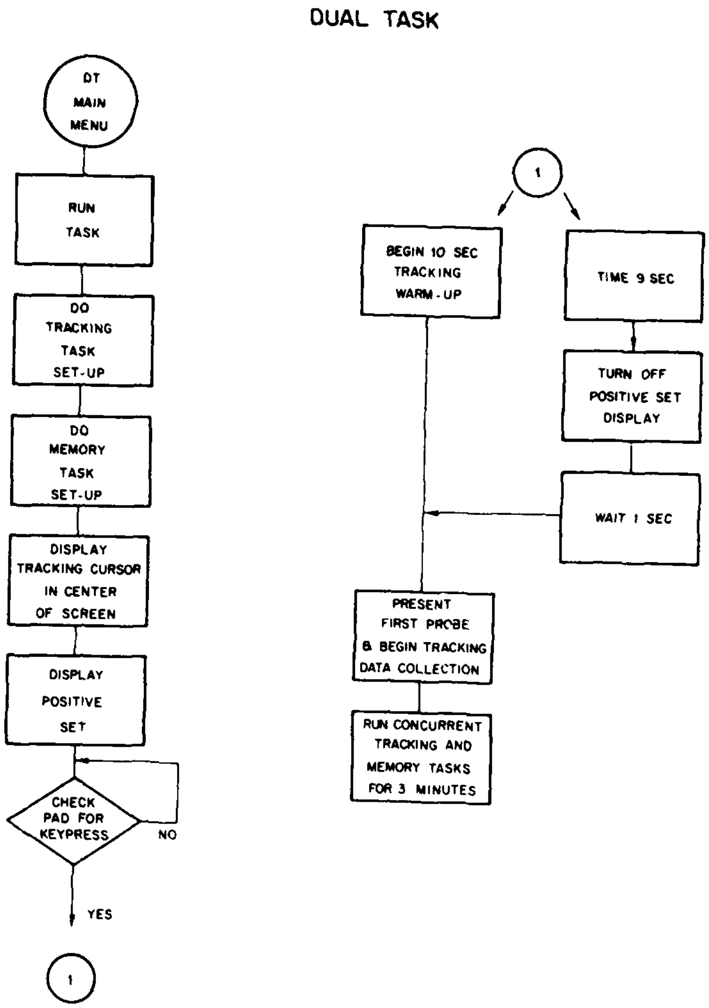


Figure 24. The structure of the Dual Task.

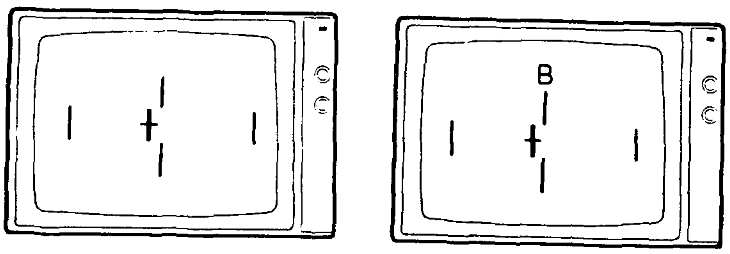


Figure 25. Example of the Dual Task display.

Training Requirements

Subjects are presented with instructions that specify that both tasks are equally important, and that the object is to respond as quickly and accurately as possible on the Memory Search task while tracking as accurately as possible.

Since each subject has previously performed the tasks in isolation, the purpose of this training phase is merely to permit practice on their concurrent performance. Initial dual-task performance is normally erratic; subjects should complete five practice blocks at each memory set size (standard schedule) or two practice blocks at each memory set size (abridged schedule). If the dual-task is administered to the subject in several experimental sessions, practice should be omitted after the first session.

Instructions to Subjects

You will now be required to perform concurrently two tasks that you have previously performed in isolation: unstable tracking and memory search. You should use your preferred hand (the hand with which you normally write) to control the joystick, and your other hand to press the response keys. The two tasks are equally important, so try not to concentrate exclusively on one at the expense of the other.

In the tracking task, your objective is to keep a cursor centred on a target area in the middle of the monitor screen. You can control the movement of the cursor by moving the joystick. Moving the stick to the right moves the cursor to the right, and moving it to the left moves the cursor to the left. The cursor initially appears on the central target but tends to move horizontally away from this position. Try to keep it centred over the target at all times. If it reaches the boundary line, it will reappear at the target position and begin moving away again. This is called a control loss and should be avoided if possible.

While you are controlling the cursor, you will be required to respond to test letters in the memory search component of the task. As before, you will be shown a 'memory set' that will contain either two or four letters, and you will be allowed to look at it for as long as you wish. When you have memorized this set, please press one of the response keys. The tracking task will then begin immediately. After a few seconds, the memory set will disappear and you will be shown a series of single test letters. As before, you must decide whether each test letter is one of the letters in the memory set. If so, press the 'yes' key; if not, press the 'no' key.

Please try to respond to the test letters as fast as you can without making any mistakes, but do not neglect the tracking task. Remember, each task is equally important.

If you do not respond to a test letter within a certain time, the next letter will appear. The memory set presented in each period will be different, so be sure to memorize it before you press the key to begin. Each period devoted to a particular memory set will last for three minutes, and will end with the message 'end of block'.

CHAPTER 3

DATA EXCHANGE

A. DATA EXCHANGE FORMAT

A standardized data format is specified to facilitate exchange of information between researchers using the STRES battery. Files should be written in ASCII code using the Data Interchange Format (DIF). Data should be stored on double sided 5.25 inch MS-DOS diskettes (40 tracks, 9 sectors) or 3.5 inch MS-DOS diskettes (80 tracks, 9 or 18 sectors); these storage media were selected because they are available to nearly all laboratories.

Each diskette should be labelled with: a) the sender's name and address; b) a brief identifier for the experimenter (see below); and c) the date of the experiment.

If more than one diskette is used to store the data from a single experiment, the diskettes should be numbered consecutively.

B. COLLECTION OF INFORMATION

The software associated with the STRES battery should include a) a routine to collect and store, prior to administration of the battery, the general information comprising Part I of the transfer file, and b) a routine to create complete transfer files in the format specified below.

C. STANDARDIZED TRANSFER FILE CONTENT

Transfer files should begin with general information that facilitates the interpretation of test results. Although brevity is desirable, the contributor is free to use as many lines as necessary.

Part I: General information

The information comprising Part I of the transfer file appears in Table 6a. A sample of a completed general information section appears in Table 6b.

Part II: Data set

Part II comprises both subject and condition information and test scores. Information concerning each subject forms a closed block that starts with the signal SOSF (Start Of Subject File). Division 1 of each block contains stable subject information such as sex and age; Division 2 contains variable information concerning the nature of the experimental condition, together with the corresponding test results.

Part II/Division 1: Subject information

Subjects are identified only by a number in the transfer file. The information appearing after SOSF, which is requested by a computer programme integrated with the task controlling software, appears in Table 7.

Part II/Division 2: Condition information and test scores

Condition information (top of Table 8) begins with the subject's session number. In the case of repeated measurement, sessions are reported in an ascending series beginning with the first session. The condition information precedes the results for each condition (Table 8) even with repeated measurements under the same experimental condition.

Overall structure of transfer file

The end of each condition data set is marked by the signal EOCD (End Of Condition Data). The end of the subject's file is marked EOSF, and is followed by the next subject's file. The complete transfer file terminates with the marker EOTF (END OF TRANSFER FILE). This arrangement is represented in Figure 26.

D. USES OF THE TRANSFER FILE

Initially, transfer files will be used for the exchange of data between individual laboratories. Eventually, however, a central data base may be established, to which users of the STRES Battery will be able to contribute and to obtain access. Such a data base, although desirable, is beyond the scope of Working Group 12's activities.

The major functions of data exchange will be a) to help to identify the psychometric properties of the tests, b) to provide normative data, c) to indicate the pattern of performance change associated with a particular stressor, d) to indicate the effects of a range of stressors on a particular mental process, e) to indicate the effects of 'incidental' variables such as age on mental performance, f) to reveal occupational differences in performance that may be of interest to selection researchers, and g) to facilitate communication between users with common interests.

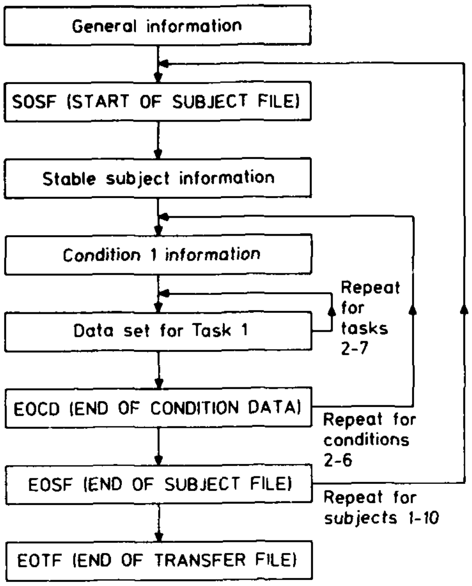


Figure 26. Structure of a transfer file for 10 subjects each completing six conditions.

Table 6a. Format of general information section of the transfer file.

Part I: General information

Title of experiment or research project:

Author(s) with address and telephone:

Short title (not more than 10 characters):

Keywords:

Summary (indicating rationale, methodology and results):

Reference (where the results are published or documented):

Date of the experiment. From: To:

Subject information: Sex:
Age range:
Education:
Occupation:
Motivation (eg payment, class credit):
Other information deemed relevant:

Number of subjects:

List of independent factors (with three letters abbreviations):
Factor 1 -- Name:
Abbreviation:
Levels of factor:
Factor 2 -- Name:
Abbreviation:
Levels of factor:
...
Factor n -- Name:
Abbreviation:
Levels of factor:

Experimental design (eg within-subjects):

Special experimental conditions:

Deviation from standard test conditions:

Table 6b. Sample general information section of data transfer file.

Part I: General information

Title of experiment or research project: Interaction of performance effects of sleep loss and noise.

Author(s): F. Smith, Dept of Psychology, University of Oxbridge, UK. Tel 456 7890

Short title: SLEEPNOIS

Keywords: Sleep loss/Noise/STRES Battery

Summary: Possible interactive effects of sleep loss and noise investigated. STRES battery administered in within-Ss design to 16 Ss in following conditions: rested/quiet; sleep-deprived/quiet; rested/noise; sleep-deprived/noise. On all tests, noise impaired the performance of rested Ss; sleep loss impaired performance under quiet conditions; but noise improved the performance of sleep-deprived Ss.

Reference: to appear in Journal of Stress Research

Date: Sept 1988 to Nov 1988

Subject information: Sex: Male.
Age range: 19-25
Education: Undergraduate
Occupation: Students
Motivation: one pound paid per session
Other information deemed relevant:

Number of subjects: 16

List of independent factors:
Factor 1 – Name: SLEEP LOSS
Abbreviation: SLL
Levels of factor: Rested, 1 night's sleep loss
Factor 2 – Name: NOISE
Abbreviation: NOI
Levels of factor: 65dB, 95dB

Experimental design: Within-Ss

Special experimental conditions: None

Deviation from standard test conditions: 10 practice blocks given on each task.

Table 7. Subject information appearing in Part II/Division 1 of the transfer file.

Subject number

Sex (m or f)

Age (years)

School education (ILL=Illiterate; BAS=Basic School Level; MED=Medium School Level; UNI=University Entrance Level)

Total years at school (including ground school)

Main occupation

Number of years in main occupation

Reported visual status (NOC = no correction necessary to view computer screen; CON = correction to normal vision, VID = Visual deficiencies, not 100% correctable)

Experience with the standardized test system (Y/N)

Special remarks (additional relevant subject characteristics)

Table 8. Test scores in Part II/Division 2 of the transfer file.

		Session number: Time of day (24 hours) Time since last session: Condition (eg sleep loss): Stressor abbreviation(s) and level(s), separated by a comma. eg NOI65dB,NOI95dB:
REACTION TIME	Basic:	1) mean RT for correct responses 2) SD of RTs for correct responses 3) number of trials 4) percent errors 5) percent response failures
		(Repeat for Coded; Time Uncertainty; Double Responses; Inversion; Basic)
MATHEMATICAL PROCESSING		1) mean RT for all correct responses 2) SD of RTs for all correct responses 3) mean RT for correct > responses 4) SD of RTs for correct > responses 5) mean RT for correct < responses 6) SD of RTs for correct < responses 7) number of > trials 8) number of < trials 9) percent errors to > problems 10) percent errors to < problems 11) percent response failures on > problems 12) percent response failures on < problems
MEMORY SEARCH	Memory set of 2	1) memory set size 2) memory search inspection time 3) mean RT for all correct responses 4) SD of RTs for all correct responses 5) mean RT for correct positive responses 6) SD of RTs for correct positive responses 7) mean RT for correct negative responses 8) SD of RTs for correct negative responses 9) number of positive trials 10) number of negative trials 11) percent errors to positive probes 12) percent errors to negative probes 13) percent response failures for pos. probes 14) percent response failures for neg. probes
		(Repeat for memory set of 4)
	For memory sets of 2 and 4	1) slope of RT function, positive probes 2) intercept of RT function, positive probes 3 slope of RT function, negative probes 4) intercept of RT function, negative probes
SPATIAL PROCESSING		1) mean RT for all correct responses 2) SD of RTs for all correct responses 3) mean RT for correct same responses 4) SD of RTs for correct same responses 5) mean RT for correct different responses 6) SD of RTs for correct different responses 7) number of same trials 8) number of different trials 9) percent errors on same trials 10) percent errors on different trials 11) percent response failures on same trials 12) percent response failures on different trials
UNSTABLE TRACKING		1) RMS error score 2) number of control losses
GRAMMATICAL REASONING		1) mean RT for all correct responses 2) SD of RTs for all correct responses 3) mean RT for correct same responses 4) SD of RTs for correct same responses 5) mean RT for correct different responses 6) SD of RTs for correct different responses 7) number of same trials 8) number of different trials 9) percent errors on same trials 10) percent errors on different trials 11) percent response failures on same trials 12) percent response failures on different trials

Table 8. (cont'd)

DUAL TASK TRACKING		1)	RMS error score
		2)	number of control losses
DUAL TASK MEMORY SEARCH	Memory Set of 2	1)	memory set size
		2)	memory search inspection time
		3)	mean RT for all correct responses
		4)	SD of RTs for all correct responses
		5)	mean RT for correct positive responses
		6)	SD of RTs for correct positive responses
		7)	mean RT for correct negative responses
		8)	SD of RTs for correct negative responses
		9)	number of positive trials
		10)	number of negative trials
		11)	percent errors to positive problems
		12)	percent errors to negative problems
		13)	percent response failures to pos. problems
		14)	percent response failures to neg. problems
			(Repeat tracking and memory search scores for memory set of 4)
	For memory sets or 2 and 4	1)	slope of RT function, positive probes
		2)	intecept of RT function, positive probes
		3)	slope of RT function, negative probes
		4	intercept of RT function, negative probes

CHAPTER 4

CONCLUSION

The goals attained by Working Group 12 during its lifespan can be summarized as follows:

1) Survey of performance researchers in NATO countries and publication of a register, as a first step in promoting exchange of information;

2) Selection of tasks for inclusion in the STRES battery;

3) Review of previous literature on each task (or on similar tasks);

4) Specification of standardized parameters for each task;

5) Specification of data exchange format.

The efforts of the working group were directed towards hardware-independent specifications, because of the wide variety of computers used by performance researchers, and the relatively short lifespan of any particular system. The refinement of the STRES Battery will take place over a protracted period, during which computer systems currently considered as industry standards may well have become obsolete. However, the introduction of high-level task development software (eg Schneider, 1988) will permit even those inexperienced in programming to construct most of the tasks specified in this report.

The objective of the STRES Battery is not to stultify performance research. The tasks selected are those that are already in common use (albeit in a variety of guises). Moreover, the yardstick provided by the battery may prove useful to those who wish to develop new approaches.

As more information concerning the psychometric properties of the tasks becomes available, the battery will evolve. It may be necessary to refine task parameters, or to introduce additional tasks.

The accumulation of data will permit some aspects of the validity of the STRES battery to be explored more fully. However, formal validation studies are also required. The working group considers the following approaches to be desirable:

1) Use of factor analysis to relate the battery to a well-established ability factor space, such as that formed by Cattell's Comprehensive Ability Factors.

2) Assessment of construct validity by administering the tests to various occupational groups. It can be predicted, for example, that a group of successful pilots will score more highly than a group of radio operators on the Spatial Processing task.

3) Assessment of the degree to which performance decrement on the tests reflects changes in real-life activity. For example, the user must be able to infer the operational consequences of a particular pattern of decrement in test scores under sleep loss.

4) Assessment of cross-cultural validity. It must be ensured that performance on the tasks is not affected by cultural differences. As discussed earlier, for example, the Grammatical Reasoning Test, as described by Baddeley (1968), would have been unsuitable for use in German, because of the avoidance of the passive voice in that language.

As performance data accumulate, it will be possible to examine more fully the reliability of each task, and the range of stressors to which it is sensitive. It will also be possible to investigate the extent to which each test is sensitive to individual differences, and the relevance of the test to occupational performance. Existing evidence concerning the psychometric properties of the STRES tasks has already been described. This evidence must, however, be considered tentative, because of variation in test procedures. The STRES battery introduces the standardization that is essential for rigorous psychometric investigation.

Further progress is dependent upon widespread use of the battery, and exchange of data between laboratories. Performance researchers interested in the STRES battery are invited to contact any member of Working Group 12 for information.

ACKNOWLEDGEMENTS

Some of the material appearing in the test descriptions was adapted from Armstrong Aerospace Medical Research Laboratory technical report AAMRL-TR-87-007 by W. A. Perez, P. J. Masline, E. G. Ramsey, and K. E. Urban, and from TNO/IZF report 1987-2 by L.C. Boer, A.W.K. Gaillard, and P.G.A.M. Jorna.

The members of the working group wish to thank Mr M. Vikmanis, AAMRL, Wright-Patterson Air Force Base, Dr J. J. Vos, TNO Institute for Perception, and Air Vice Marshal P. Howard, RAF Institute of Aviation Medicine, for their encouragement and information.

REFERENCES

Adam, N., & Collins, G. Late components of the visual evoked potential to search in short term memory. Electroencephalography and Clinical Neurophysiology, 1978, 44, 147-156.

Adler, A., Strasser, H., & Muller-Limmroth, W. Effects of massed and distributed practice, additional monetary and placebo incentives, and changes in characteristics of control element during learning of human motor control. Perceptual and Motor Skills, 1976, 42, 1335-1338.

AGARD Aerospace Medical Panel Working Group 12. Performance Assessment Register. Neuilly-sur-Seine: AGARD Report No. 763, 1988.

Anders, T. R., Fozard, J. L., & Lillyquist, T. D. (1972). Effects of age upon retrieval from short term memory. Developmental Psychology, 1972, 6, 214-217.

Ashcraft, M. H., & Battaglia, J. Cognitive arithmetic: Evidence for retrieval and decision processes in mental addition. Journal of Experimental Psychology: Human Learning and Memory, 1978, 4, 527-538.

Ashcraft, M. H., & Stazyk, E. H. Mental addition: A test of three verification models. Memory and Cognition, 1981, 9, 185-196.

Ashkenas, I. L., & McRuer, D. T. The Determination of Material Handling Quality Requirements from Airframe/Human-Pilot System Studies. Wright-Patterson Air Force Base, Ohio: Wright Air Development Center Report No. TR 59-135, 1959.

Baddeley, A. D. A 3-minute reasoning test based on grammatical transformations. Psychonomic Science, 1968, 10, 341-342.

Baddeley, A. D. Working Memory. Oxford: Oxford University Press, 1986.

Baddeley, A. D., de Figueredo, J. W., Hawkswell Curtis, J. W., & Williams, A. N. Nitrogen narcosis and underwater performance. Ergonomics, 1968, 11, 157-164.

Baddeley, A. D., & Hitch, G. J. Working memory. In G. H. Bower (Ed.), The Psychology of Learning and Motivation Volume 8. New York: Academic Press, 1974.

Boer, L. C. [Psychological Fitness of Leopard 1-V Crews After a 200-km Drive.] In Dutch. Soesterberg, Netherlands: TNO Institute for Perception Report No. IZF 1987-30, 1987.

Boer, L. C. [A Normative Database for, and Analysis of, Some Taskomat Tasks.] In Dutch. Soesterberg, Netherlands: TNO Institute for Perception, Report No. IZF 1988-10, 1988.

Boer, L. C., Gaillard, A. W. K., & Jorna, P. G. A. M. [Taskomat, A Task Battery for Information Processing.] In Dutch. Soesterberg, Netherlands: TNO Institute for Perception, Report No. IZF 1982-2.

Boer, L. C., Ruzius, M. H. B., Mimpen, A. M., Bles, W., & Janssen, W. H. Psychological Fitness During a Manoeuvre. Soesterberg, Netherlands: TNO Institute for Perception Report No. IZF 1984-17, 1984.

Briggs, G. E., & Blaha, J. Memory retrieval and central comparison times in information processing. Journal of Experimental Psychology, 1969, 79, 395-402.

Briggs, G. E., & Johnsen, A. M. On the nature of central processing in choice reactions. Memory and Cognition, 1973, 1, 91-100.

Briggs, G. E., Peters, G. L., & Fisher, R. P. On the locus of the divided-attention effects. Perception and Psychophysics, 1972, 11, 315-320.

Briggs, G. E., & Swanson, J. M. Retrieval time as a function of memory ensemble size. Quarterly Journal of Experimental Psychology, 1969, 21, 185-191.

Brookhuis, K. A., Mulder, G., Mulder, L. J. M., Gloerich, A. B. M., van Dellen, H. J., van der Meere, J. J., & Ellermann, H. Late positive components and stimulus evaluation time. Biological Psychology, 1981, 13, 107-123.

Brown, I. D., Tickner, A. H., & Simmonds, D. C. V. Interference between concurrent tasks of driving and telephoning. Journal of Applied Psychology, 1969, 53, 419-424.

Burrows, D., & Okada, R. Serial position effects in high-speed memory search. Perception and Psychophysics, 1971, 10, 305-308.

Burrows, D., & Okada, R. Parallel scanning of semantic and formal information. Journal of Experimental Psychology, 1973, 97, 254-257.

Carter, R. C., Kennedy, R. S., & Bittner, A. C. Grammatical reasoning: A stable performance yardstick. Human Factors, 1981, 23, 587-591.

Carter, R. C., Kennedy, R. S., Bittner, A. C., & Krause, M. Item recognition as a performance evaluation test for environmental research. Proceedings of the 24th Annual Meeting of the Human Factors Society, Santa Monica, 1980, 340-344.

Carter, R. C., & Krause, M. Reliability of slope scores for individuals: Experiment 4: Letter search. New Orleans: Naval Dynamics Laboratory Report No. NBDL-83R003, 1983, 17-20.

Chase, W. G., & Calfee, R. C. Modality and similarity effects in short term recognition memory. Journal of Experimental Psychology, 1969, 81, 510-514.

Chase, W. G., & Clark, H. H. Mental operations in the comparison of sentences and pictures. In L. W. Gregg (Ed.), Cognition in Learning and Memory. New York: Wiley and Sons, 1972.

Chiles, W. D., & Alluisi, E. A. On the specification of operator or occupational workload with performance-measurement methods. Human Factors, 1979, 21, 515-528.

Chiles, W. D., Alluisi, E. A., & Adams, O. S. Work schedules and performance during confinement. Human Factors, 1968, 10, 143-196.

Chiles, W. D., Bruni, C. B., & Lewis, R. A. Methodology in the Assessment of Complex Human Performance: The Effects of Signal Rate on Monitoring a Dynamic Process. Washington, D.C.: U.S. Department of Transportation, Federal Aviation Administration Report No. FAA-AM-69-9, 1969.

Chiles, W. D., & Jennings, A. E. Effects of alcohol on complex performance. Human Factors, 1970, 12, 605-612.

Chiles, W. D., Jennings, A. E., & Alluisi, E. A. The Measurement and Scaling of Workload in Complex Performance. Washington, D.C.: U.S. Department of Transportation, Federal Aviation Administration Report No. FAA-AM-7834, 1978.

Clark, H. H., & Chase, W. G. On the process of comparing sentences against pictures. Cognitive Psychology, 1972, 3, 472-517.

Clark, H. H., & Chase, W. G. Perceptual coding strategies in the formation and verification of descriptions. Memory and Cognition, 1974, 2, 101-111.

Clifton, C., & Tash, J. Effect of syllabic word length on memory search rate. Journal of Experimental Psychology, 1973, 99, 231-235.

Connor, J. M. Serial and parallel encoding processes in memory and visual search. Journal of Experimental Psychology, 1972, 96, 363-370.

Cooper, L. A., & Shepard, R. N. Transformations on representations of objects in space. In E. C. Carterette and M. P. Friedman (Eds.), Handbook of Perception, Vol 8: Space and Object Perception. New York: Academic Press, 1978.

Crosby, J. V., & Parkinson, G. R. A dual task investigation of pilots' skill level. Ergonomics, 1979, 22, 1301-1313.

Damos, D. L., Bittner, A. C., Kennedy, R. S., & Harbeson, M. M. Effects of extended practice on dual task tracking performance. Human Factors, 1981, 23, 627-631.

Damos, D. L., Bittner, A. C., Kennedy, R. S., Harbeson, M. M., & Krause, M. K. Performance evaluation tests for environmental research (PETER): Critical tracking test. Perceptual and Motor Skills, 1984, 58, 567-573.

DeRosa, D. V. Transformation on sets in short term memory: Set size reduction by deletion. Journal of Experimental Psychology, 1969, 82, 415-426.

DeRosa, D. V., & Morin, R. E. Recognition reaction time for digits in consecutive and nonconsecutive sets. Journal of Experimental Psychology, 1970, 83, 472-479.

DeRosa, D. V., & Sabol, M. Transformations on sets in short term memory: Temporal and spatial factors influencing deletion. Memory and Cognition, 1973, 1, 69-72.

Donders, F.C. [On the speed of mental processes.] In W.G. Koster (Ed. and Trans.), Attention and performance II, 412-431. Amsterdam: North Holland, 1969. (Reprinted from Acta Psychologica, 1969, 30.)

Dott, A. B., & McKelvy, R. K. Influence of ethyl alcohol in moderate levels on visual stimulus tracking. Human Factors, 1977, 19, 191-199.

Englund, C. E., Reeves, D. L., Shingledecker, C. A., Thorne, D. R., Wilson, K. P., & Hegge, F. W. Unified Tri-service Cognitive Performance Assessment Battery (UTC-PAB). 1: Design and Specification of the Battery. San Diego, California: Naval Health Research Center Report No. 87-10, 1987.

Eysenck, M. W. & Eysenck, M. C. Memory scanning, introversion-extraversion and levels of processing. Journal of Personality Research, 1979, 13, 305-315.

Farmer, E. W., Berman, J. V. F., & Fletcher, Y. L. Evidence for a visuo-spatial scratch-pad in working memory. Quarterly Journal of Experimental Psychology, 1986, 38A, 675-688.

Farmer, E. W., & Green, R. G. The sleep-deprived pilot: Performance and EEG response. In M. Sorsa (Ed.), Report of the Sixteenth Conference of the Western European Association for Aviation Psychology. Helsinki: Finnair Training Centre, 1985, 155-162.

Fitts, P. M., Weinstein, M., Rappaport, M., Anderson, N., & Leonard, J. A. Stimulus correlates of visual pattern perception: A probability approach. Journal of Experimental Psychology, 1956, 51, 1-11.

Ford, J. M., Roth, W. T., Mohs, R. C., Hopkins, W. F., & Kopell, B. S. Event related potentials recorded from young and old adults during a memory retrieval task. Electroencephalography and Clinical Neurophysiology, 1979, 47, 450-459.

Foss, D. J., & Dowell, B. E. High-speed memory retrieval with auditorily presented stimuli. Perception and Psychophysics, 1971, 9, 465-468.

Frowein, H. W. Selective Drug Effects on Information Processing. Dissertatie, Katholieke Hogeschool Tilburg, 1981.

Frowein, H. W., Gaillard, A. W. K., & Varey, C. A. EP components, visual processing stages, and the effect of a barbiturate. Biological Psychology, 1981, 13, 239-249.

Frowein, H. W., Reitsma, D., & Aquarius, C. Effects of two counteracting stresses on the reaction process. In J. Long and A. D. Baddeley (Eds.), Attention and Performance IX. Hillsdale, New Jersey: Erlbaum, 1981.

Gaillard, A. W. K., Gruisen, A., & de Jong, R. The Influence of Loratadine (Sch 29851) on Human Performance. Soesterberg, Netherlands: TNO Institute for Perception Report No. IZF 1986-C19, 1986.

Gaillard, A. W. K., Rozendaal, A. H., & Varey, C. A. [The Effects of Marginal Vitamin-Deficiency on Mental Performance.] In Dutch. Soesterberg, Netherlands: TNO Institute for Perception Report No. IZF 1983-29, 1983.

Gaillard, A. W. K., Varey, C. A., & Ruzius, M. H. B. Marginal Vitamin Deficiency and Mental Performance. Soesterberg, Netherlands: TNO Institute for Perception Report No. IZF 1985-22, 1985.

Gaillard, A. W. K., & Verduin, C. J. The combined effects of an antihistamine and pseudo-ephedrine on human performance. Journal of Drug Research, 1983, 8, 1929-1936.

Gomer, F.E., Spicuzza, R. J., & O'Donnell, R. D. Evoked potential correlates of visual item recognition during memory scanning tasks. Physiological Psychology, 1976, 4, 61-65.

Graboi, D. Searching for targets: The effects of specific practice. Perception and Psychophysics, 1971, 10, 300-304.

Greenewald, A. G. Time-Sharing, Ideomotor Compatibility, and Automaticity. Proceedings of the 23rd Annual Meeting of the Human Factors Society, 1979.

Grether, W. F. Acceleration and Human Performance. Wright-Patterson Air Force Base, Ohio: Aerospace Medical Research Laboratory Report No. AMRL-TR-71-22, 1971.

Hall, T. J., Passey, G. E., & Meighan, T. W. Performance of Vigilance and Monitoring Tasks as a Function of Workload. Wright-Patterson Air Force Base, Ohio: Aerospace Medical Research Laboratories Report No. AMRL-TR-65-22, 1965.

Hann, R. L. The Sternberg paradigm and memory scan research. Major Area Paper, Vanderbilt University, 1973.

Harris, G. J. & Fleer, R. E. High-speed memory scanning in mental retardates: Evidence for a central processing deficit. Journal of Experimental Child Psychology, 1974, 17, 452-459.

Hitch, G. J. The role of short term working memory in mental arithmetic. Cognitive Psychology, 1978, 10, 302-323.

Hitch, G. J., & Baddeley, A. D. Verbal reasoning and working memory. Quarterly Journal of Experimental Psychology, 1976, 28, 603-621.

Hockey, R., & Hamilton, P. The cognitive patterning of stress states. In R. Hockey (Ed.), Stress and Fatigue in Human Performance. Chichester: John Wiley & Sons, 1983.

Holland, P., Kemp, K. H., & Wetherell, A. Some effects of 2 mg im atropine and 5 mg im diazepam, separa ·ly and combined, on human performance. British Journal of Clinical Phamacology, 1978, 5, 367P.

Houghton, J. O., McBride, D. K., & Hannah, K. Performance and physiological effects of acceleration-induced (+Gz) loss of consciousness. Aviation, Space, and Environmental Medicine, 1985, 56, 956-965.

Hoving, K. L., Morin, R. E., & Konick, D. S. Recognition reaction time and size of the memory set: A developmental study. Psychonomic Science, 1970, 21, 247-248.

Hunt, E. Intelligence as an information-processing concept. British Journal of Psychology, 1980, 71, 449-474.

Jex, H. R., & Cromwell, C. H. Theoretical and Experimental Investigation of Some New Longitudinal Handling Quality Parameters. Wright-Patterson Air Force Base, Ohio: Aeronautical Systems Division Report No. TR 61-26, 1961.

Jex, H. R., McDonnell, J. D., & Phatak, A. V. A "critical" tracking task for manual control research. IEEE Transactions on Human Factors in Electronics, 1966, HFE-7, 138-144.

Jex, H. R., Peters, R. A., DiMarco, R. J., & Allen, R. W. The Effects of Bedrest on Crew Performance During Simulated Shuttle Reentry, Vol II: Control Task Performance. Washington, D.C.: NASA Contractor Report No. NASA CR 2367, 1974.

Keele, S. W., & Hawkins, H. L. Explorations of individual differences relevant to high skill level. Journal of Motor Behaviour, 1982, 14, 3-23.

Kemp, K. H. & Wetherell, A. Some Effects of a Single 10mg Oral Dose of Diazepam on the Psychomotor Performance of Normal Men. Porton Down, Wilts: Chemical Defence Establishment Report No. TN 326, 1977.

Kennedy, R. S., Dunlap, W. P. Jones, M. B., Lane, N. E., & Wilkes, R. L. Portable Human Assessment Battery: Stability, Reliability, Factor Structure, and Correlation with Tests of Intelligence. Final Technical Report, National Science Foundation, 1985.

Klatzky, R. L., & Atkinson, R. C. Specialization of the hemispheres in scanning for information in short term memory. Perception and Psychophysics, 1971, 10, 335-338.

Klatzky, R. L., Juola, J. F., & Atkinson, R. C. Test stimulus representation and experimental context effects in memory scanning. Journal of Experimental Psychology, 1971, 87, 281-288.

Klein, R., & Armitage, R. Rhythms in human performance: 1 1/2 hour oscillations in cognitive style. Science, 1979, 204, 1326-1328.

Klein, R., & Jex, H. R. Effects of alcohol on a critical tracking task. Journal of Studies on Alcohol, 1975, 36, 11-20.

Kristofferson, M. W. When item recognition and visual search functions are similar. Perception and Psychophysics, 1972, 12, 379-384. (a)

Kristofferson, M. W. Effects of practice on character classification performance. Canadian Journal of Psychology, 1972, 26, 540-560. (b)

Leonard, J. A. Five-Choice Serial Reaction Apparatus. Cambridge: Medical Research Council Applied Psychology Unit Report No. 326/59, 1959.

Lewis, V. J., & Baddeley, A. D. Cognitive performance, sleep quality, and mood during deep oxyhelium diving. Ergonomics, 1981, 24, 773-793.

Little, V. Z., Hartman, B. O., & Leverett, S. D. Effects of Acceleration on Human Performance and Physiology with Special Reference to Transverse G. Brooks Air Force Base, Texas: USAF School of Aerospace Medicine, Aerospace Medical Division (AFSC) Aeromedical Review No. 4-68, 1968.

Lively, B. L. Speed/accuracy tradeoff and practice as determinants of stage duration in a memory-search task. Journal of Experimental Psychology, 1972, 96, 97-103.

Lively, B. L., & Sanford, B. J. The use of category information in a memory search task. Journal of Experimental Psychology, 1972, 93, 379-385.

Logie, R. H., & Baddeley, A. D. A trimix saturation dive to 660m: Studies of cognitive performance, mood, and sleep quality. Ergonomics, 1983, 26, 359-374.

Logsdon, R., Hochhaus, L., Williams, H. L., Rundell, O. H., & Maxwell, D. Secobarbital and perceptual processing. Acta Psychologica, 1984, 55, 179-193.

Lohman, D. F. Spatial Ability: A review and Reanalysis of the Correlational Literature. Stanford, California: Aptitude Research Project, School of Education, Stanford University, Technical Report No. 8, 1979.

Lorenz, B., & Lorenz, J. Cognitive performance monitoring during the TITAN VIII and GUSI *11* simulated saturation dives: Experiences with a microcomputer-based test battery. Paper presented at Fourteenth Annual Scientific Meeting of the European Undersea Biomedical Society, Aberdeen, Scotland, 1988.

Maizlish, N. A., Langolf, G. D., Whitehead, L. W., Fine, L. J., Albers, J. W, Goldberg, J., & Smith, P. Behavioural evaluation of workers exposed to mixtures of organic solvents. British Journal of Industrial Medicine, 1985, 42, 579-590.

Massaro, D. W. Experimental Psychology and Information Processing. Chicago: Rand McNally, 1975.

McNair, D. M., Kahn, R. J., Frankenthaler, L. M., & Faldetta, L. L. Amoxapine and amitriptyline II: Specificity of cognitive effects during brief treatment of depression. Psychopharmacology, 1984, 83, 134-139.

McRuer, D. T., Graham, D., Krendel, E., & Reisner, W. Human Pilot Dynamics in Compensatory Systems: Theory, Models and Experiments with Controlled Element and Forcing Function Variations. Wright-Patterson Air Force Base, Ohio: US Air Force Flight Dynamics Laboratory Report No. TR-65-15, 1965.

McRuer, D. T., & Jex, H. R. A review of quasi-linear pilot models. IEEE *Transactions on Human Factors in Electronics*, 1967, 8, 231-249.

Michon, J. A. Tapping regularity as a measure of perceptual motor load. Ergonomics, 1966, 9, 401-412.

Miller, G. A., Galanter, E., & Pribram, K. H. Plans and the Structure of Behavior. New York: Holt, Rinehart, Winston, 1960

Mohs, R. C., Tinklenberg, J. R., Roth, W. Y., & Kopell, B. S. Sensitivity of some human cognitive functions to effects of methamphetamine and secobarbital. Drug and Alcohol Dependence, 1980, 5, 145-150.

Moraal, J. Age and Information Processing: An Application of Sternberg's Additive Factor Method. Soesterberg, Netherlands: TNO Institute for Perception Report No. IZF 1982-18, 1982.

Morin, R. E., DeRosa, D. V., & Stultz, V. Recognition memory and reaction time. Acta Psychologica, 1967, 27, 298-305.

Naus, M. J. Memory search of categorized lists: A consideration of alternative self-terminating search strategies. Journal of Experimental Psychology, 1974, 102, 992-1000.

Naus, M. J., Glucksberg, S., & Ornstein, P. A. Taxonomic word categories and memory search. Cognitive Psychology, 1972, 3, 643-654.

Naylor, H., Halliday, R., & Callaway, E. The effect of methylphenidate on information processing. Psychopharmacology, 1985, 86, 90-95.

Oborne, D. J., & Rogers, Y. Interactions of alcohol and caffeine on human reaction time. Aviation, Space, and Environmental Medicine, 1983, 54, 528-534.

Orenstein, H. B., & Hamilton, K. M. Memory load, critical features and retrieval processes in facial recognition. Perception and Motor Skills, 1977, 45, 1079-1087.

Perez, W. A. Mental Arithmetic: The Processing and Maintenance of Information in Working Memory. Unpublished Doctoral Dissertation, Miami University, Oxford, Ohio, 1982.

Pfefferbaum, A., Ford, J. M., Roth, W. T., & Kopell, B. S. Age differences in P300 reaction time associations. Electroencephalography and Clinical Neurophysiology, 1980, 49, 257-265.

Posner, M. I., & Mitchell, R. F. *Chronometric analysis of classification*. Psychological Review, 1967, 74, 392-401.

Poulton, E. C., and Freeman, P. R. Unwanted asymmetrical transfer effects with balanced experimental designs. Psychological Bulletin, 1966, 66, 1-8.

Putz, V. R. The effects of carbon monoxide on dual-task performance. Human Factors, 1979, 21, 13-24.

Putz, V. R., Anderson, V., Setzer, J. V., & Croxton, J. S. Effects of alcohol, caffeine, and methyl chloride on man. Psychological Reports, 1981, 48, 715-725.

Putz, V. R., Johnson, B. L., & Setzer, J. V. A comparative study of the effects of carbon monoxide and methylene chloride on human performance. Journal of Environmental Pathology and Toxicology, 1979, 2, 97-112.

Repko, J. D., Jones, P. E., Garcia, L. S., Schneider, E. J., Roseman, E., & Corum, C. R. Behavioral and Neurological Effects of Methyl Chloride. Washington, D.C.: U.S. Government Printing Office, U.S. DHHS (NIOSH) Publication No. 77-125, 1976.

Restle, F. Speed of adding and comparing numbers. Journal of Experimental Psychology, 1970, 83, 274-278.

Rizzuto, A. P. Diazepam and its Effects on Psychophysiological and Behavioral Measures of Performance. Wright-Patterson Air Force Base, Ohio: Armstrong Aerospace Medical Research Laboratory Report No. AAMRL-TR-87-O74. 1987.

Rizzuto, A. P., Wilson, G. F., Yates, R. E., & Palmer, R. Diazepam and its Effects on Psychophysiological Measures of Performance. Wright-Patterson Air Force Base, Ohio: Armstrong Aerospace Medical Research Laboratory Report No. AFAMRL-TR-85-O36, 1985.

Ross, J. Extended practice with a single-character classification task. Perception and Psychophysics, 1970, 8, 276-278.

Roth, W. T., Tinklenberg, J. R., & Kopell, B. S. Ethanol and marihuana effects on event-related potentials in a memory retrieval paradigm. Electroencephalography and Clinical Neurophysiology, 1977, 42, 381-388.

Rundell, O. H., Williams, H. L., & Lester, B. K. Secobarbital and information processing. Perceptual and Motor Skills, 1978, 46, 1255-1264.

Sanders, A. F. Towards a model of stress and human performance. Acta Psychologica, 1983, 53, 61-97.

Sanders, A. F., Haygood, R. C., Schroiff, H.-W., & Wauschkuhn, C. H. Standardization of Performance Tests: A Proposal for Further Steps. Aachen: Institute for Psychology, RWTH, Scientific Report Phase I (1 Sept 1985-31 Aug 1986) prepared for USAF Air Force Office of Scientific Research and European Office of Aerospace Research and Development, 1986.

Sanders, A. F., Wijnen, J. L. C., & van Arkel, A. E. An additive factor analysis of the effects of sleep-loss on reaction processes. Acta Psychologica, 1982, 51, 41-59.

Schlegel, R. E., & Gilliland, K. Evaluation of the Criterion Task Set. Wright-Patterson Air Force Base, Ohio. Armstrong Aerospace Medical Research Laboratory Report No. AAMRL-TR-87-???, in press.

Schneider, W. Micro Experimental Laboratory: An integrated system for IBM PC compatibles. Behavioral Research Methods, Instrumentation, and Computers, 1988, 2O, 2O6-217.

Schneider, W., & Shiffrin, R. M. Controlled and automatic human information processing: I. Detection, search, and attention. Psychological Review, 1977, 84, 1-66.

Seales, D. M., Kennedy, R. S., & Bittner, A. C. Development of performance evaluation tests for environmental research (PETER): Arithmetic computation. Perceptual and Motor Skills, 1980, 51, 1023-1031.

Shiffrin R. M., & Schneider, W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 1977, 84, 127-190.

Shingledecker, C. A. A Task Battery for Applied Human Performance Assessment Research. Wright-Patterson Air Force Base, Ohio: Air Force Aerospace Medical Research Laboratories Report No. AFAMRL-TR-84-071, 1984.

Shingledecker, C. A., Acton, W. H., & Crabtree, M. S. Development and application of a criterion task set for workload metric evaluation. Paper presented at the Second Aerospace Behavioral Engineering Technology Conference, (No. 83-419), Long Beach, California, 1983.

Slobin, D. I. Grammatical transformations and sentence comprehension in childhood and adulthood. Journal of Verbal Learning and Verbal Behaviour, 1966, 5, 219-227.

Smith, P. J., & Langolf, G. D. The use of Sternberg's memory scanning paradigm in assessing effects of chemical exposure. Human Factors, 1981, 23, 701-708.

Stazyk, E. H., Ashcraft, M. H., & Hamann, M. S. A network approach to mental multiplicati m. Journal of Experimental Psychology: Learning, Memory, and Cognition, 1982, 8, 320-335.

Sternberg, S. High-speed scanning in human memory. Science, 1966, 153, 652-654.

Sternberg, S. Two operations in character recognition: Some evidence from reaction time measurements. Perception and Psychophysics, 1967, 2, 45-53.

Sternberg, S. Memory scanning: Mental processes revealed by reaction time experiments. American Scientist, 1969, 57, 421-457. (a)

Sternberg, S. The discovery of processing stages: Extensions of Donders' method. In W.G. Koster (Ed.), Attention and performance II, 276-315. Amsterdam: North Holland, 1969. (Reprinted from Acta Psychologica, 1969, 30). (b)

Sternberg, S. Decomposing mental processes with reaction time data. Invited address to Midwestern Psychological Association, Detroit, 1971.

Sternberg, S. Memory scanning: New findings and current controversies. Quarterly Journal of Experimental Psychology, 1975, 27, 1-32.

Steyvers, F. J. J. M. The influence of sleep deprivation and knowledge of results on perceptual encoding. Acta Psychologica, 1987, 66, 173-178.

Stokx, L. C., and Gaillard, A. W. K. Task and driving performance of patients with a severe concussion of the brain. Journal of Clinical and Experimental Neuropsychology, 1986, 8, 421-436.

Subhan, Z. The effects of benzodiazepines on short term memory and information processing. Psychopharmacology, 1984, 1, 173-181.

Sverko, B. Individual Differences in Time-Sharing Performance. Champaign, Illinois: University of Illinois, Savoy Aviation Research Laboratory Report No. ARL-77-4/ AFOSR-77-4, 1977.

Swanson, J. M., Johnsen, A. M., & Briggs, G. E. Recoding in a memory search task. Journal of Experimental Psychology, 1972, 93, 1-9.

Tharp, V. K., Rundell, O. H., Lester, B. K., & Williams, H. L. Alcohol and information processing. Psychopharmacology, 1974, 40, 33-52.

Theios, J., Smith, P. G., Haviland, S. E., Traupmann, J., & Moy, M. C. Memory scanning as a serial self-terminating process. Journal of Experimental Psychology, 1973, 97, 323-336.

Ussher, M. H., & Farmer, E. W. Anxiety prior to and during decompression. In E. D. Megaw (Ed.), Contemporary Ergonomics 1987. London: Taylor and Francis, 1987.

Van Patten, R. E. Tolerance, fatigue, physiological and performance effects of sustained and oscillating lateral acceleration. In AGARD Conference Proceedings No. 371 (Human Factors Considerations in High Performance Aircraft), 1984.

Vidulich, M., & Wickens, C. D. Time-Sharing, Manual Control and Memory Search: The Joint Effects of Input and Output Modality Competition, Priorities and Control Order. Champaign, Illinois: Department of Psychology, University of Illinois, Report No. EPL-81-4, 1981.

Wanner, E. and Shiner, S. Measuring transient memory load. Journal of Verbal Learning and Verbal Behavior, 1976, 15, 159-167.

Ward, M. M., Sandman, C. A., George, J. M., & Shulman, H. MSH Melanocyte stimulating hormone ACTH 4-10 in men and women: Effects upon performance of an attention and memory task. Physiology and Behaviour, 1979, 22, 669-674.

Wason, P. C. Response to affirmative and negative binary statements. British Journal of Psychology, 1961, 52, 133-142.

Welford, A. T. (Ed.) Reaction Times. London: Academic Press, 1980.

Wetherell, A. Studies on a Test of Higher Mental Function. Porton Down, Wilts: Chemical Defence Establishment Report No. TN 278, 1976.

Wetherell, A. The efficacy of some auditory-vocal subsidiary tasks as measures of the mental load on male and female drivers. Ergonomics, 1981, 24, 197-214.

Wetherell, A. Some factors affecting spatial memory for route information. In R. S. Easterby & H. J. G. Zwaga (eds.), Information Design. Chichester: Wiley, 1984. (a)

Wetherell, A. Physostigmine and Memory. Porton Down, Wilts: Chemical Defence Establishment Report No. TN 634, 1984. (b)

Wetherell, A. Physostigmine and Information Processing in Memory. Porton Down, Wilts: Chemical Defence Establishment Report No. 696, 1986

Wickens, C. D. The effects of divided attention on information processing in manual tracking. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 1-13.

Wickens, C. D. Processing Resources in Attention, Dual Task Performance, and Workload Assessment. Champaign, Illinois: University of Illinois Engineering Psychology Research Laboratory, Technical Report No. EPL-81-3, 1981.

Wickens, C. D. Engineering Psychology and Human Performance. Columbus, Ohio: Charles E. Merrill Publishing Co., 1984.

Wickens, C. D., Mountford, S. J., & Schreiner, W. Task Dependent Differences and Individual Differences in Dual Task Performance. New Orleans: Naval Biodynamics Laboratory Report No. NBDL-M003, 1980.

Wickens, C. D., & Sandry, D. Task-hemispheric integrity in dual task performance. Acta Psychologica, 1982, 52, 227-247.

Wickens, C. D., Sandry, D. L., & Hightower, R. Display Location of Verbal and Spatial Material: The Joint Effects of Task-Hemispheric Integrity and Processing Strategy. Champaign, Illinois: University of Illinois, Department of Psychology, Report No. EPL-82-2, 1982.

Wickens, C. D., Vidulich, M., Sandy, D., & Schiflett, S. Factors influencing the performance advantage of speech technology. Proceedings of the 25th Annual Meeting of the Human Factors Society, Rochester, NY, 1981.

Williams, H. L., Rundell, O. H., & Smith, L. T. Dose effects of secobarbital in a Sternberg memory scanning task. Psychopharmacology, 1981, 72, 161-165.

Wingfield, A., & Branca, A. A. Strategy in high-speed memory search. Journal of Experimental Psychology, 1970, 83, 63-67.

Yerkes, R. M., & Dodson, J. D. The relation of strength of stimulus to rapidity of habit-formation. Journal of Comparative and Neurological Psychology, 1908, 18, 459-482.

ANNEX 1

PROJECT SUMMARY

Human performance assessment is used by all NATO countries in their aerospace programmes. Human operator performance should be measured or estimated in all systems using humans. This includes measuring the effectiveness of new systems, measuring operator workload, determining the effects of environmental stressors, and assisting in the design of new systems. Despite the widespread use of performance tests, there are few, if any, accepted testing methods that are the same in all laboratories; hence, it is very difficult to share results between laboratories and countries.

Human performance assessment is an important problem, especially since systems are becoming more complex and demanding of the operator. Since the overall effectiveness of these systems depends upon the human, it is crucial that performance assessment be accurate. The ability to share data would permit much more rapid progress in the design and construction of military systems that best utilized the operator's capabilities.

The purpose of Working Group 12 is to establish the methodology by which standardized tests will be selected. This includes not only the selection of tests but also the specification of their parameters, identification of the areas in which they are useful, and provision of a data exchange format and a bibliography. A core test battery will be determined so that the use of these accepted tests can begin. To achieve its goals, the Working Group will undertake the following activities:

I. Compilation of a register of performance assessment laboratories and personnel. The draft of this document will be completed by June 1987 with the USAF responsible for its collation.

II. Preparation of an AGARDograph on recommended tests for stress testing and performance assessment. The procedures used for selection of these tests will also be discussed in the AGARDograph. A draft of this document will be completed by January 1988.

III. A Lecture Series will be made available to member countries. Countries not represented in the Working Group will be especially targeted for this series, which will be held during 1988 following the distribution of the AGARDograph.

IV. An AGARD Symposium will be proposed in 1989 in conjunction with an AMP meeting. The proceedings of this international meeting on 'Human Performance Assessment Methods' will be published.

The first meeting of the Working Group will be held at Wright-Patterson Air Force Base, Ohio, USA during the latter part of January 1987. It is suggested that subsequent meetings will be held at TNO Institute for Perception, The Netherlands, the RAF Institute of Aviation Medicine, Great Britain, and DFVLR, West Germany. These meetings will be held at six-monthly intervals for the duration of the two-year term of the Working Group.

The Working Group will also interact with the so called 'Academic Group'. The Academic Group had its first meeting at Aachen, West Germany, in the fall of 1984 and had a second meeting at Paris, France in the spring of 1986. These meetings were funded by EOARD in London. The Working Group will solicit input from the Academic Group. The Working Group has definite goals and a two-year life, and so it must accomplish its goals within this period. The Academic Group is concerned more with theoretical discussions and test development, whereas the Working Group is concerned with applications of tests in military environments. Both groups are interested in performance assessment and share some members; appropriate information from the Academic Group is thus available to the Working Group.

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's Reference	3. Further Reference	4. Security Classification of Document
	AGARD-AG-308	ISBN 92-835-0510-7	UNCLASSIFIED

5. Originator Advisory Group for Aerospace Research and Development
North Atlantic Treaty Organization
7 rue Ancelle, 92200 Neuilly sur Seine, France

6. Title HUMAN PERFORMANCE ASSESSMENT METHODS

7. Presented on 5–6 June 1989 in Downsview (Toronto), Canada, on 12–13 June 1989 in Soesterberg, The Netherlands and on 15–16 June 1989 in Pratica di Mare (Rome), Italy.

8. Author(s)/Editor(s)	9. Date
Various	May 1989

10. Author's/Editor's Address	11. Pages
Various	70

12. Distribution Statement This document is distributed in accordance with AGARD policies and regulations, which are outlined on the Outside Back Covers of all AGARD publications.

13. Keywords/Descriptors

Performance evaluation	Tests
Psychometrics	Surveys
STRES battery	Data base

14. Abstract

AGARDograph 308 presents the results of the second phase of AGARD Aerospace Medical Panel Working Group 12 on "Human Performance Assessment Methods". The major goal of WG 12 was to develop the "Standardized Tests for Research on Environmental Stressors" or "STRES" Battery, satisfying conventional psychometric criteria such as reliability, validity and sensitivity for which an extensive data base may now be compiled among the NATO nations. The protocol for the 7 selected tests is presented.

AGARDograph 308 also serves as the Lecture Series Notes for AGARD Aerospace Medical Panel Lecture Series 163, "Human Performance Assessment Methods". This Lecture Series has been implemented by the AGARD Consultant and Exchange Programme.

AGARDograph No.308
Advisory Group for Aerospace Research and Development, NATO
HUMAN PERFORMANCE ASSESSMENT METHODS
Published May 1989
72 pages

AGARDograph 308 presents the results of the second phase of AGARD Aerospace Medical Panel Working Group 12 on "Human Performance Assessment Methods". The major goal of WG 12 was to develop the "Standardized Tests for Research on Environmental Stressors" or "STRES" Battery, satisfying conventional psychometric criteria such as reliability, validity and sensitivity for which an extensive data base may now be compiled among the NATO nations. The protocol for the

P.T.O.

AGARD-AG-308

Performance evaluation
Psychometrics
STRES battery
Tests
Surveys
Data base

AGARDograph No.308
Advisory *Group for Aerospace Research* and Development, NATO
HUMAN PERFORMANCE ASSESSMENT METHODS
Published May 1989
72 pages

AGARDograph 308 presents the results of the second phase of AGARD Aerospace Medical Panel Working Group 12 on "Human Performance Assessment Methods". The major goal of WG 12 was to develop the "Standardized Tests for Research on Environmental Stressors" or "STRES" Battery, satisfying conventional psychometric criteria such as reliability, validity and sensitivity for which an extensive data base may now be compiled among the NATO nations. The protocol for the

P.T.O.

AGARD-AG-308

Performance evaluation
Psychometrics
STRES battery
Tests
Surveys
Data base

AGARDograph No.308
Advisory Group for Aerospace Research and Development, NATO
HUMAN PERFORMANCE ASSESSMENT METHODS
Published May 1989
72 pages

AGARDograph 308 presents the results of the second phase of AGARD Aerospace Medical Panel Working Group 12 on "Human Performance Assessment Methods". The major goal of WG 12 was to develop the "Standardized Tests for Research on Environmental Stressors" or "STRES" Battery, satisfying conventional psychometric criteria such as reliability, validity and sensitivity for which an extensive data base may now be compiled among the NATO nations. The protocol for the

P.T.O.

AGARD-AG-308

Performance evaluation
Psychometrics
STRES battery
Tests
Surveys
Data base

AGARDograph No.308
Advisory *Group for Aerospace Research* and Development, NATO
HUMAN PERFORMANCE ASSESSMENT METHODS
Published May 1989
72 pages

AGARDograph 308 presents the results of the second phase of AGARD Aerospace Medical Panel Working Group 12 on "Human Performance Assessment Methods". The major goal of WG 12 was to develop the "Standardized Tests for Research on Environmental Stressors" or "STRES" Battery, satisfying conventional psychometric criteria such as reliability, validity and sensitivity for which an extensive data base may now be compiled among the NATO nations. The protocol for the

P.T.O.

AGARD-AG-308

Performance evaluation
Psychometrics
STRES battery
Tests
Surveys
Data base

7 selected tests is presented.

AGARDograph 308 also serves as the Lecture Series Notes for AGARD Aerospace Medical Panel Lecture Series 163, "Human Performance Assessment Methods". This Lecture Series has been implemented by the AGARD Consultant and Exchange Programme.

ISBN 92-835-0510-7

7 selected tests is presented.

AGARDograph 308 also serves as the Lecture Series Notes for AGARD Aerospace Medical Panel Lecture Series 163, "Human Performance Assessment Methods". This Lecture Series has been implemented by the AGARD Consultant and Exchange Programme.

ISBN 92-835-0510-7

7 selected tests is presented.

AGARDograph 308 also serves as the Lecture Series Notes for AGARD Aerospace Medical Panel Lecture Series 163, "Human Performance Assessment Methods". This Lecture Series has been implemented by the AGARD Consultant and Exchange Programme.

ISBN 92-835-0510-7

7 selected tests is presented.

AGARDograph 308 also serves as the Lecture Series Notes for AGARD Aerospace Medical Panel Lecture Series 163, "Human Performance Assessment Methods". This Lecture Series has been implemented by the AGARD Consultant and Exchange Programme.

ISBN 92-835-0510-7

AGARD

NATO OTAN

7 rue Ancelle · 92200 NEUILLY-SUR-SEINE

FRANCE

Telephone (1)47.38.57.00 · Telex 610 176

DISTRIBUTION OF UNCLASSIFIED AGARD PUBLICATIONS

AGARD does NOT hold stocks of AGARD publications at the above address for general distribution. Initial distribution of AGARD publications is made to AGARD Member Nations through the following National Distribution Centres. Further copies are sometimes available from these Centres, but if not may be purchased in Microfiche or Photocopy form from the Purchase Agencies listed below.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordonnateur AGARD — VSL
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA
Director
Dept of N
Ottawa, C

DENMARK
Danish D
Ved Idrae
2100 Cop

FRANCE
O.N.E.R.
29 Avenu
92320 Ch

GERMANY
Fachinfor
Physik, M
Karlsruhe
D-7514 E

GREECE
Hellenic A
Aircraft Su
Departmen
Holargos, Athens, TGA 1010

ICELAND
Director of Aviation
c/o Flugrad
Reyjavik

ITALY
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
3 Piazzale Adenauer
00144 Roma/EUR

LUXEMBOURG
See Belgium

NETHERLANDS
Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR

NASA

National Aeronautics and Space Administration

Washington, D.C. 20546

SPECIAL FOURTH CLASS MAIL BOOK

Postage and Fees Paid
National Aeronautics and Space Administration
NASA-451

Official Business
Penalty for Private Use $300

L1 001 AGARDAG3088906145002672D
DEPT OF DEFENSE
DEFENSE TECHNICAL INFORMATION CENTER
DTIC-FDAC
CAMERON STATION BLDG 5
ALEXANDRIA VA 223046145

ent

ARD

Ankara

UNITED KINGDOM
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

UNITED STATES
National Aeronautics and Space Administration (NASA)
Langley Research Center
M/S 180
Hampton, Virginia 23665

THE UNITED STATES NATIONAL DISTRIBUTION CENTRE (NASA) DOES NOT HOLD STOCKS OF AGARD PUBLICATIONS, AND APPLICATIONS FOR COPIES SHOULD BE MADE DIRECT TO THE NATIONAL TECHNICAL INFORMATION SERVICE (NTIS) AT THE ADDRESS BELOW.

PURCHASE AGENCIES

National Technical
Information Service (NTIS)
5285 Port Royal Road
Springfield
Virginia 22161, USA

ESA/Information Retrieval Service
European Space Agency
10, rue Mario Nikis
75015 Paris, France

The British Library
Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
England

Requests for microfiche or photocopies of AGARD documents should include the AGARD serial number, title, author or editor, and publication date. Requests to NTIS should include the NASA accession report number. Full bibliographical references and abstracts of AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
published by NASA Scientific and Technical Information Branch
NASA Headquarters (NIT-40)
Washington D.C. 20546, USA

Government Reports Announcements (GRA)
published by the National Technical Information Services, Springfield
Virginia 22161, USA

Printed by Specialised Printing Services Limited
40 Chigwell Lane, Loughton, Essex IG10 3TZ

ISBN 92-835-0510-7